

# GRADIENT BOOST

**Селезнёв Артём**

Аналитик больших данных  
2019



# Градиентный бустинг скрывает в себе

1

$$Set(x_i, x_j)_{i=1}^n$$

и функцию потерь

$$L(y_i, F(x))$$

2

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$

3

Вычисление

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

4

Обучение регрессии

$$r_{jm}$$

в цикле  $j = 1 \dots m$

5

В цикле  $j = 1 \dots m$  вычисление

$$r_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

6

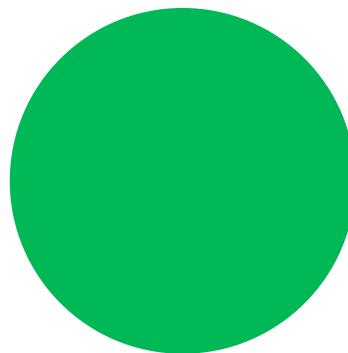
$$F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$$

Разберем  
на примере.  
Что скрыто  
в ARGMIN?

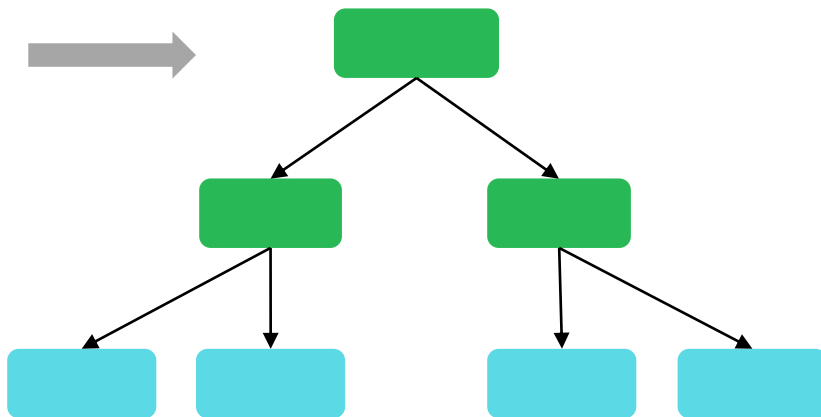
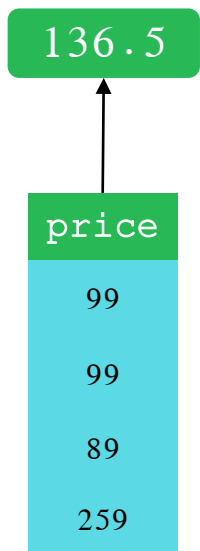
# Наш помощник – набор данных

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

# Вспомните деревья



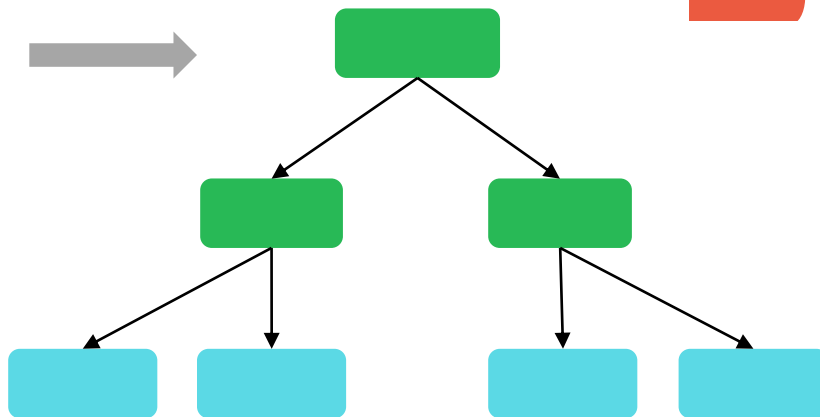
## Вспомните деревья



## Вспомните деревья

136.5

price
99
99
89
259

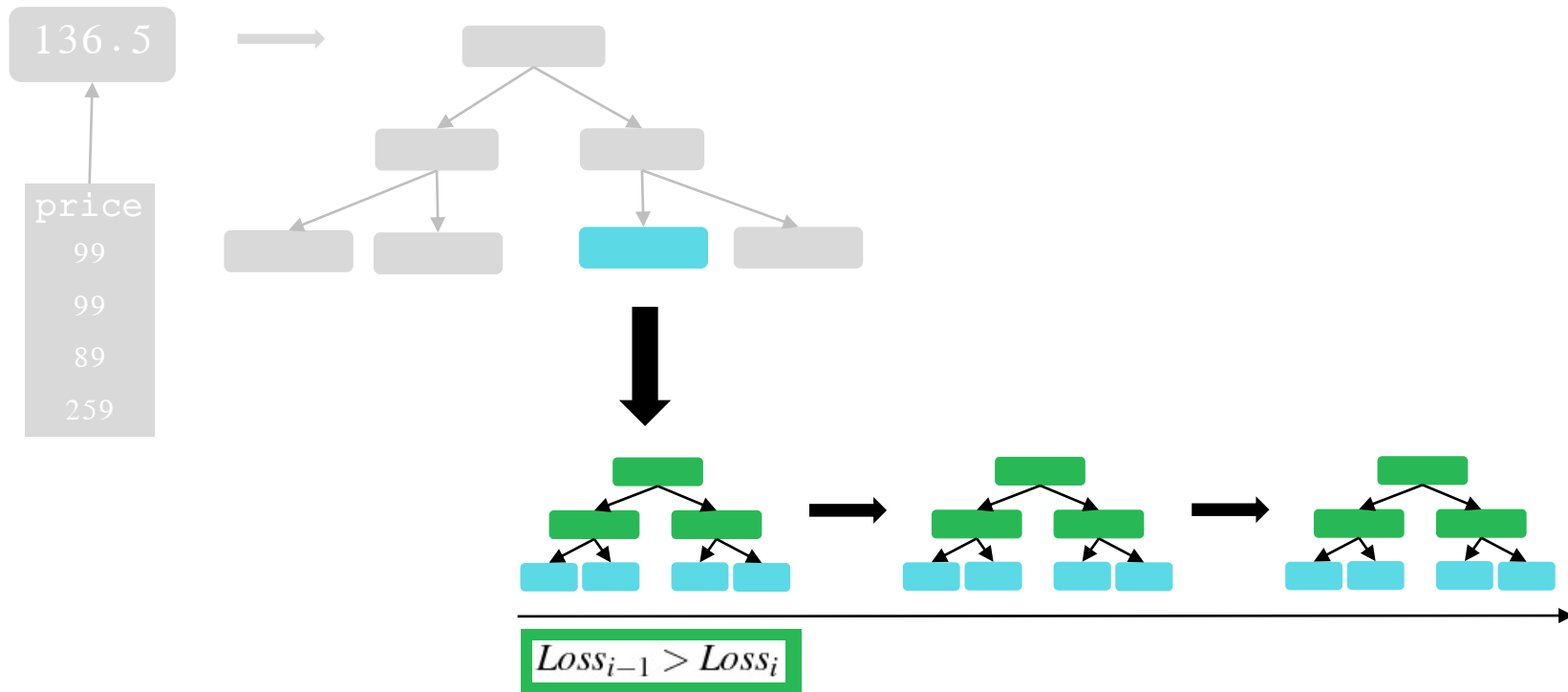


Глубина должна  
быть ограничена.

обычно 8 – 32



## Вспомните деревья





# Начинаем считать

price
99
99
89
259

→ 136,5

$$\begin{aligned} & (\text{Observed} - \text{Predicted}) \\ &= (99 - 136.5) = -37.5 \end{aligned}$$

Region	Category	Sub category	Param1	threshold	price	diff
Владимирская область	Личные вещи	1	1	3	99	
Волгоградская область	Личные вещи	2	1	2	99	
Кировская область	Личные вещи	5	1	3	89	
Москва	Личные вещи	4	6	3	259	



# Начинаем считать

price
99
99
89
259

→ 136,5

$$\begin{aligned} & (\text{Observed} - \text{Predicted}) \\ &= (99 - 136.5) = -37.5 \end{aligned}$$

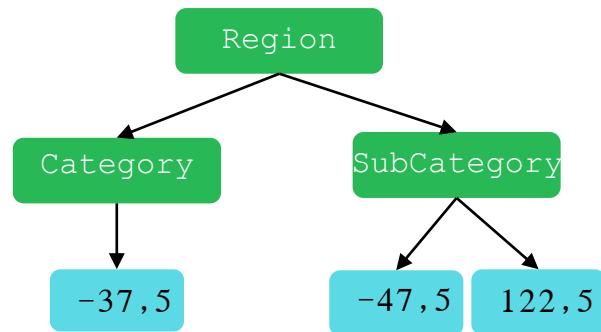


Region	Category	Sub category	Param1	threshold	price	diff
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



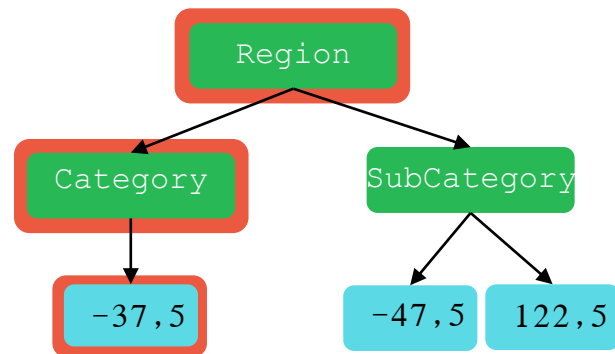
# Делаем первое дерево

Region	Category	Subcategory	Param1	threshold	price	T
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



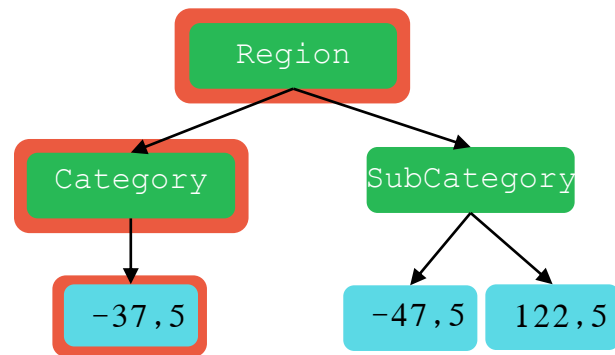
# Делаем первое дерево

Region	Category	Subcategory	Param1	threshold	price	T
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



# Делаем первое дерево

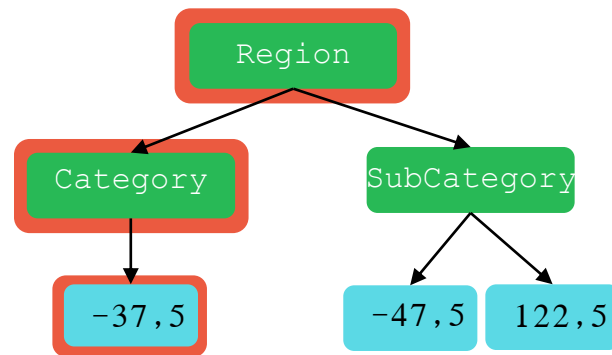
Region	Category	Subcategory	Param1	threshold	price	T
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



$$\begin{aligned} & (\text{Avg Price} + \text{Predicted}) \\ &= (136.5 + (-37.5)) = 99 \end{aligned}$$

# Делаем первое дерево

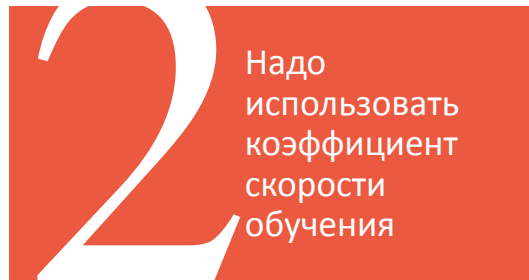
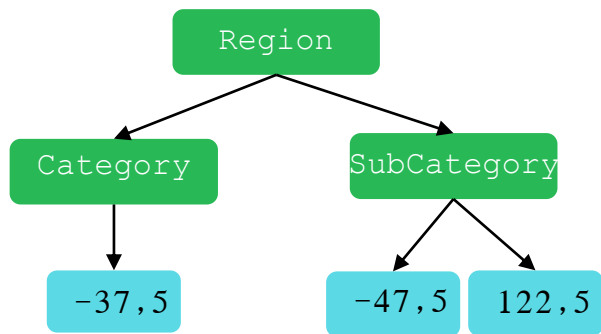
Region	Category	Subcategory	Param1	threshold	price	T
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



$$\begin{aligned} & (\text{Avg Price} + \text{Predicted}) \\ & = (136.5 + (-37.5)) = \end{aligned}$$

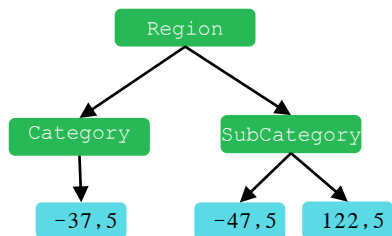
99

# Делаем первое дерево



$$\begin{aligned} & (\text{Avg Price} + (\text{Learning Rate} * \text{Predicted})) \\ & = 136.5 + (0.1 * (-37.5)) = 132.75 \end{aligned}$$

# Делаем первое дерево



$$\left( \text{Avg Price} + \left( \text{Learning Rate} * \text{Predicted} \right) \right)$$

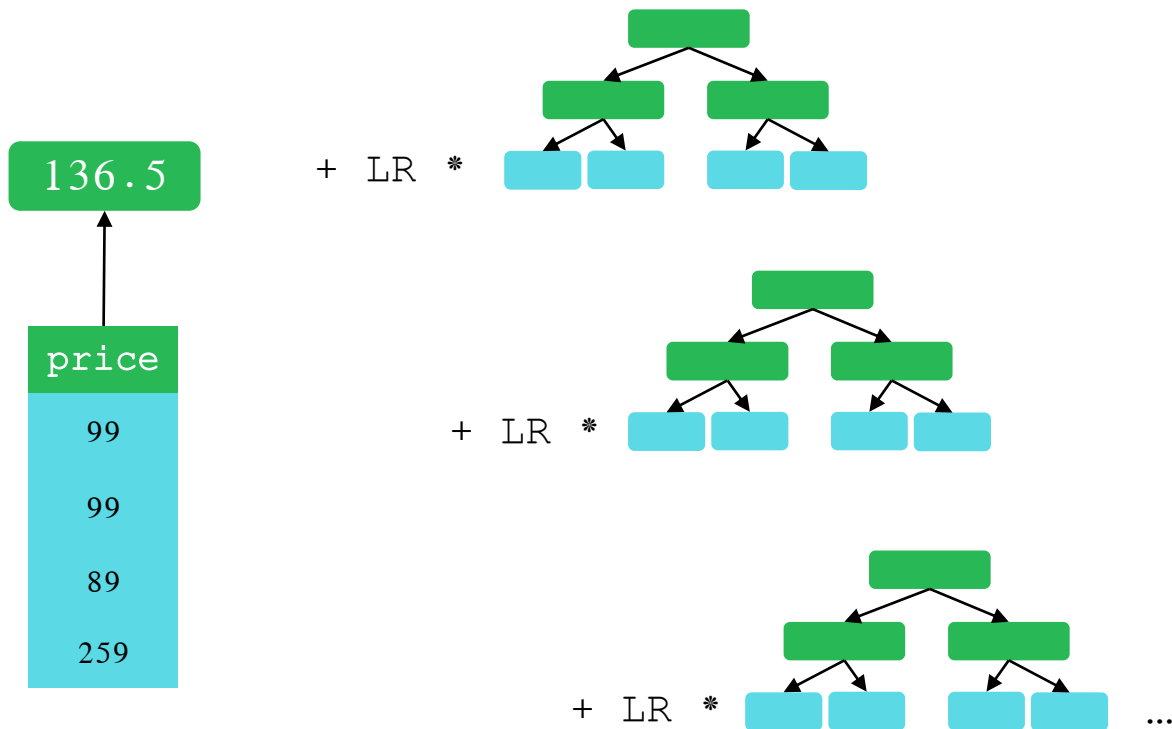


Region	Category	Subcategory	Param1	threshold	price	diff
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5





# Ещё больше деревьев



\*LR - learning  
rate



**Смотрите**

# Мы разобрали:



1  $Set(x_i, x_j)_{i=1}^n$  и функцию потерь  $L(y_i, F(x))$

2  $F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$

3 Вычисление  $r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$

4 Обучение регрессии  $r_{jm}$  в цикле  $j = 1 \dots m$

5 В цикле  $j = 1 \dots m$  вычисление  $\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$

6  $F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$

**Давайте  
закрепим**

# Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^n$$

$$Set(x_i, x_j)_{i=1}^n$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$L(y_i, F(x))$$



(Observed – Predicted)

# Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^n$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$Set(x_i, x_j)_{i=1}^n$$

$$L(y_i, F(x))$$



$$\frac{diff}{diff * Predicted} * \frac{1}{2} (Observed - Predicted)^2 = \frac{1}{2} (Observed - Predicted)^2$$

# Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^n$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$Set(x_i, x_j)_{i=1}^n$$

$$L(y_i, F(x))$$



$$\frac{diff}{diff * Predicted} * \frac{1}{2} (Observed - Predicted)^2 = \frac{1}{2} (Observed - Predicted)^2$$

# Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^n$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$Set(x_i, x_j)_{i=1}^n$$

$$L(y_i, F(x))$$



$$\frac{1}{2}(Observed - Predicted)^2 = \frac{2}{2}(Observed - Predicted)x - 1 = -(Observed - Predicted)$$



# Наш помощник – набор данных

$$Set(x_i, x_j)_{i=1}^n$$

Region	Category	Subcategory	Param1	threshold	price
Владимирская область	Личные вещи	1	1	3	99
Волгоградская область	Личные вещи	2	1	2	99
Кировская область	Личные вещи	5	1	3	89
Москва	Личные вещи	4	6	3	259

$$Set(x_i, x_j)_{i=1}^n$$

$$L(y_i, F(x))$$



$$\frac{1}{2}(Observed - Predicted)^2 = \frac{1}{2}(Observed - Predicted)x - 1 = -(Observed - Predicted)$$

Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$



Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$



price
99
99
89
259



$$\frac{1}{2}(99 - \text{Predicted})^2 + \frac{1}{2}(99 - \text{Predicted})^2 + \frac{1}{2}(89 - \text{Predicted})^2 + \frac{1}{2}(259 - \text{Predicted})^2$$

# Argmin <- минимизируем потерю



$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$



price
99
99
89
259

$$\frac{1}{2}(99 - \text{Predicted})^2 \rightarrow -(99 - \text{Predicted}) +$$

$$\frac{1}{2}(99 - \text{Predicted})^2 \rightarrow -(99 - \text{Predicted}) +$$

$$\frac{1}{2}(89 - \text{Predicted})^2 \rightarrow -(89 - \text{Predicted}) +$$

$$\frac{1}{2}(259 - \text{Predicted})^2 \rightarrow -(259 - \text{Predicted})$$

Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$



$$- 99 + \text{Predicted} + (-99) + \text{Predicted} + (-89) + \text{Predicted} + (-259) + \text{Predicted} = 0$$

price
99
99
89
259



# Argmin <- минимизируем потерю



$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$



$$- 99 + \text{Predicted} + (-99) + \text{Predicted} + (-89) + \\ \text{Predicted} + (-259) + \text{Predicted} = 0$$



$$\text{Predicted} + \text{Predicted} + \text{Predicted} \\ = 99 + 89 + 259$$

# Argmin <- минимизируем потерю



$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$



$$- 99 + \text{Predicted} + (-99) + \text{Predicted} + (-89) + \text{Predicted} + (-259) + \text{Predicted} = 0$$



$$4 * \text{Predicted} = (2 * 99) + 89 + 259$$



$$\frac{(2 * 99) + 89 + 259}{4}$$

# 2

## Градиентный спуск



# Теперь разберём :



1  $Set(x_i, x_j)_{i=1}^n$  и функцию потерь  $L(y_i, F(x))$

2  $F_0(x) = argmin \Sigma L(y_i, \gamma)$

3 Вычисление  $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$

4 Обучение регрессии  $r_{jm}$  в цикле  $j = 1...m$

5 В цикле  $j = 1...m$  вычисление  $r_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$

6  $F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$

**Сначала  
посмотрим**

# Разбираем шаги спуска

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$



# Разбираем шаги спуска

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$



$$\frac{diff}{diff * Predicted} * \frac{1}{2} (Observed - Predicted)^2$$



# Разбираем шаги спуска

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$



# Разбираем шаги спуска

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$



(Observed  $F_{m-1}(x)$ )



(Observed  $F_0(x)$  |  $F_0(x)$ ) → 136.



# Помните этот слайд?

price
99
99
89
259

→ 136,5

$$\begin{aligned} & (\text{Observed} - \text{Predicted}) \\ &= (99 - 136.5) = -37.5 \end{aligned}$$



Region	Category	Sub category	Param1	threshold	price	diff
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5

# Он считает R

price
99
99
89
259

→ 136,5

$$\begin{aligned} & (\text{Observed} - \text{Predicted}) \\ &= (99 - 136.5) = -37.5 \end{aligned}$$



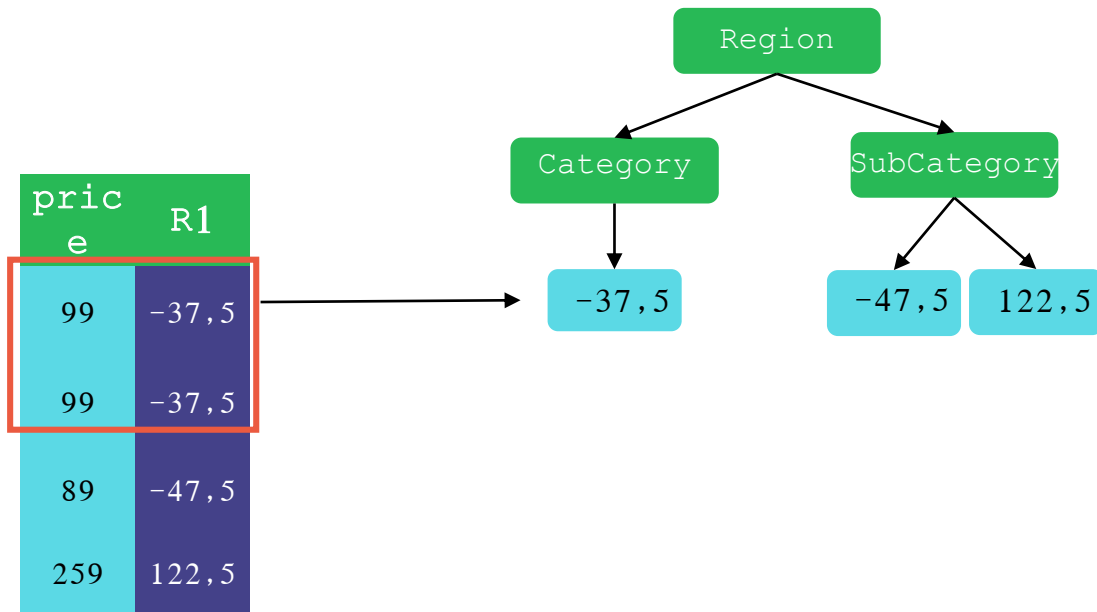
Region	Category	Sub category	Param1	threshold	price	R1
Владимирская область	Личные вещи	1	1	3	99	-37,5
Волгоградская область	Личные вещи	2	1	2	99	-37,5
Кировская область	Личные вещи	5	1	3	89	-47,5
Москва	Личные вещи	4	6	3	259	122,5



# Обучение регрессии

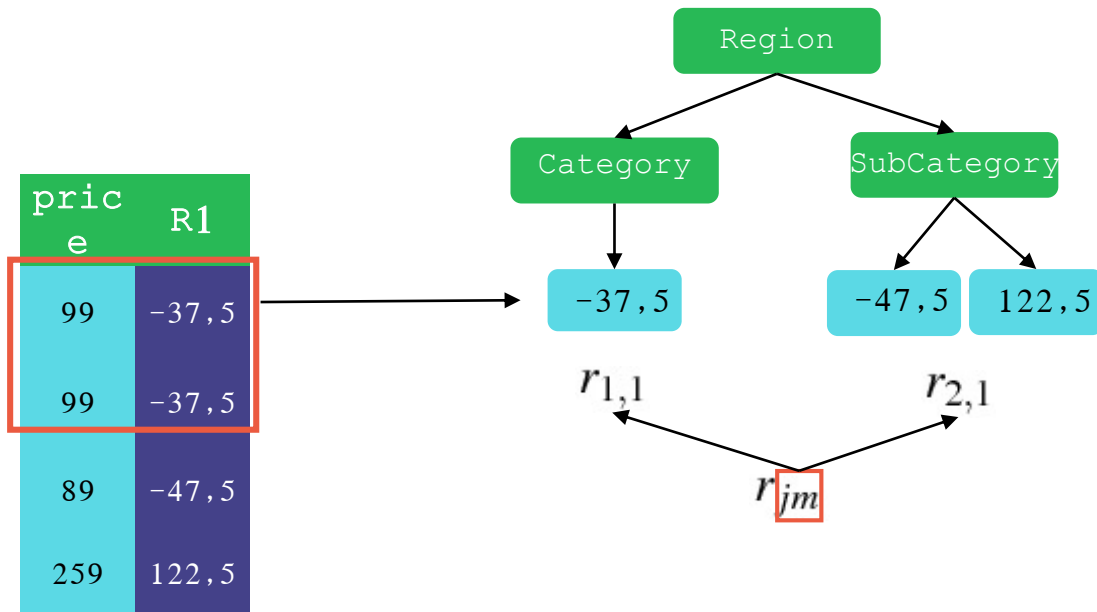
в цикле  $j = 1..m$

$r_{jm}$



# Обучение регрессии в цикле $j = 1..m$

$r_{jm}$

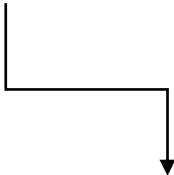


В цикле  $j = 1 \dots m$   
вычисление  $\gamma$   $\operatorname{argmin}$

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

В цикле  $j = 1 \dots m$   
вычисление  $\gamma$   $\operatorname{argmin}$

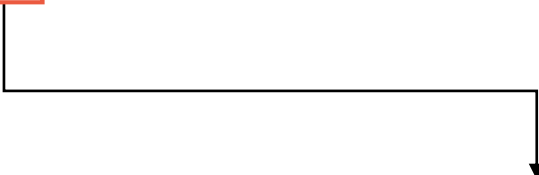
$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$

В цикле  $j = 1 \dots m$

вычисление  $\gamma$   $\text{argmin}$

$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



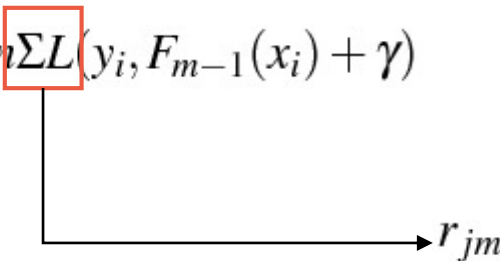
A diagram consisting of a horizontal line with a vertical segment at the left end and a downward-pointing arrow at the right end, connecting the boxed  $\gamma_{jm}$  in the first equation to the boxed  $\gamma$  in the second equation.

$$F_0(x) = \text{argmin} \Sigma L(y_i, \gamma)$$

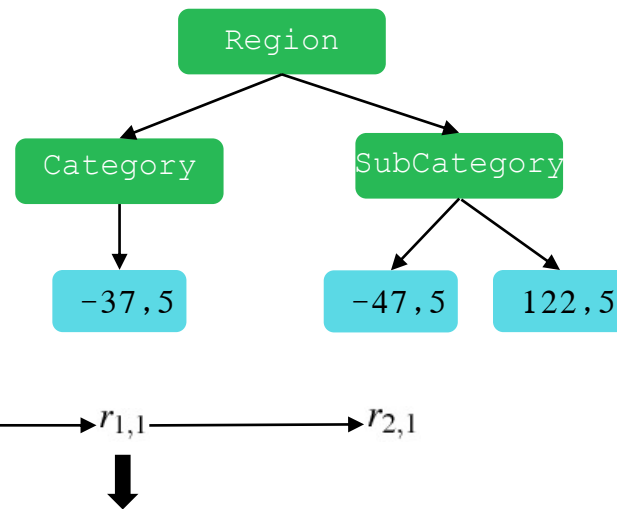
В цикле  $j = 1 \dots m$   
вычисление  $\gamma_{jm}$   $\operatorname{argmin}$

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

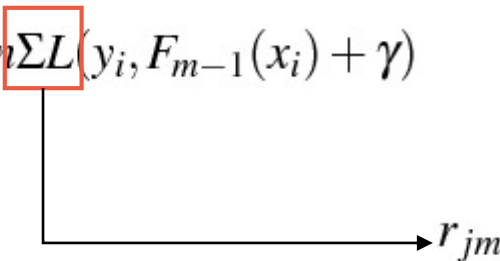
В цикле  $j = 1 \dots m$   
вычисление  $\gamma$   $\text{argmin}$

$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


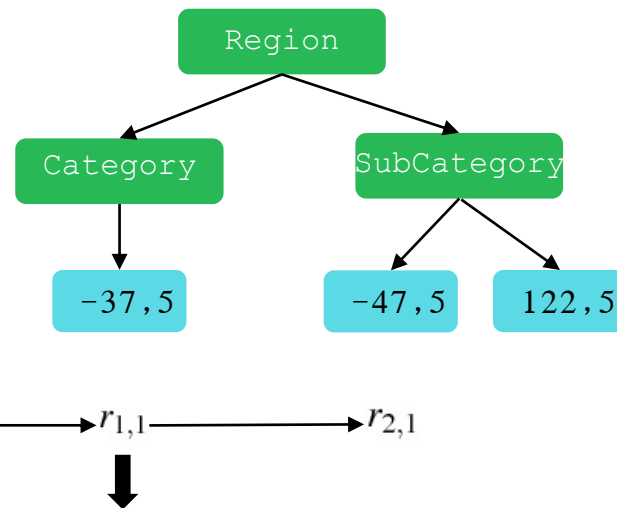
A diagram showing the flow from the boxed 'argmin' in the equation to the variable  $r_{jm}$ . A vertical line descends from the box, then a horizontal line extends to the right, ending in an arrow pointing to  $r_{jm}$ .

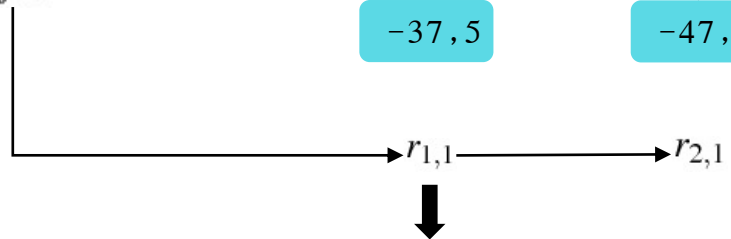


В цикле  $j = 1 \dots m$   
вычисление  $\gamma$   $\text{argmin}$

$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


A diagram showing a line from the boxed  $\text{argmin}$  in the equation above, leading to the variable  $r_{jm}$ .



$$\gamma_{1,1} = \text{argmin} \frac{1}{2} (y_i - (F_{m-1}(x_i) + \gamma))^2 = \text{argmin} \frac{1}{2} (-37.5 - \gamma)^2$$


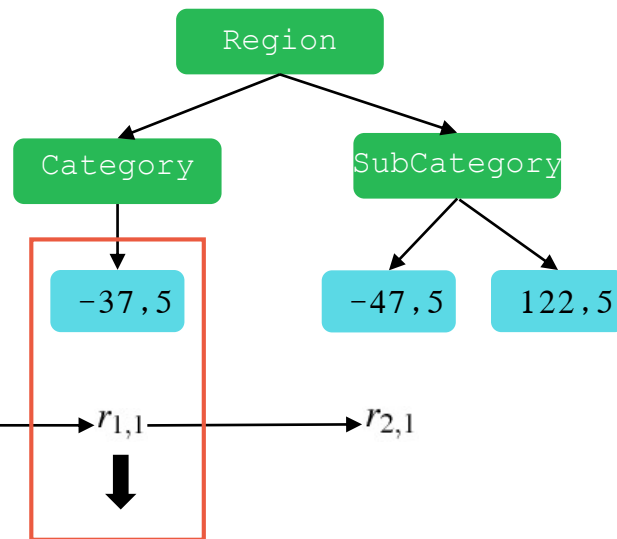
A diagram showing a line from  $r_{1,1}$  (which is the value -37.5 from the tree) leading to the final equation. A thick downward arrow is also present below  $r_{1,1}$ .



В цикле  $j = 1 \dots m$   
 вычисление  $\gamma$   $\text{argmin}$

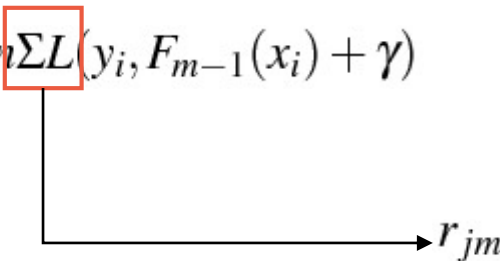
$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$r_{jm}$

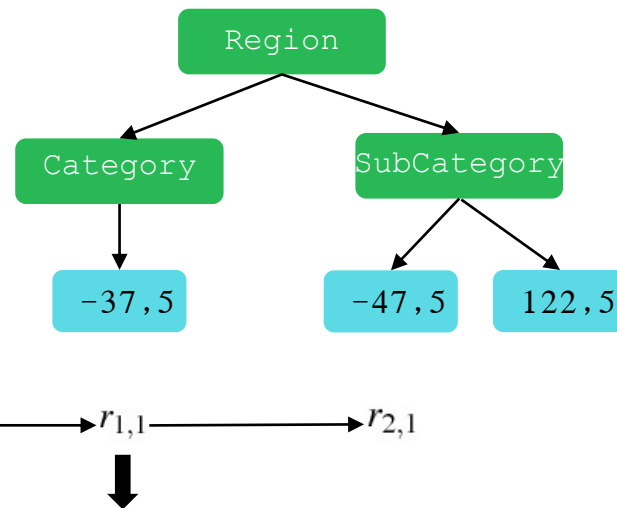


$$\gamma_{1,1} = \text{argmin} \frac{1}{2} (y_i - (F_{m-1}(x_i) + \gamma))^2 = \text{argmin} \frac{1}{2} (-37.5 - \gamma)^2$$

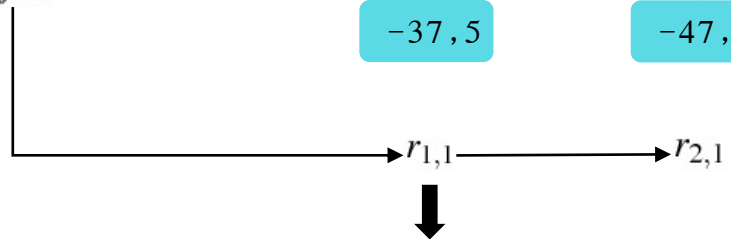
В цикле  $j = 1 \dots m$   
вычисление  $\gamma$   $\text{argmin}$

$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


A diagram showing a line from the boxed  $\text{argmin}$  in the equation above, extending downwards and then to the right to point at the variable  $r_{jm}$ .



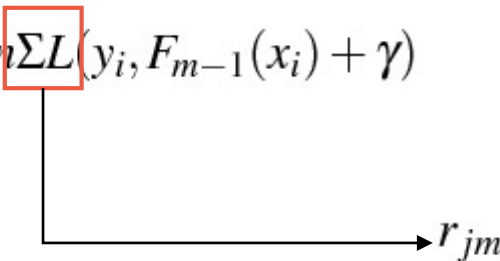
$r_{1,1}$   $\rightarrow$   $r_{2,1}$



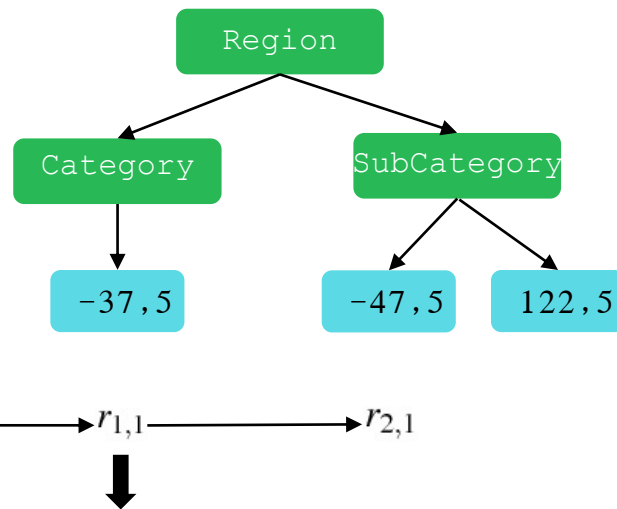
A diagram showing a line from  $r_{1,1}$  extending to the right to point at  $r_{2,1}$ . Below  $r_{1,1}$ , a thick downward arrow points to the final equation.

$$\gamma_{1,1} = \text{argmin} \frac{1}{2} (y_i - (F_{m-1}(x_i) + \gamma))^2 = \text{argmin} \frac{1}{2} (-37.5 - \gamma)^2$$

В цикле  $j = 1 \dots m$   
вычисление  $\gamma$   $\text{argmin}$

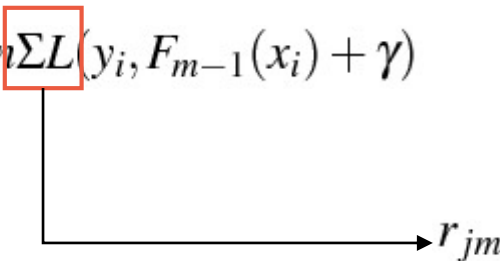
$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


A diagram showing a line from the boxed  $\Sigma$  in the equation above, going down and then right to point at the variable  $r_{jm}$ .

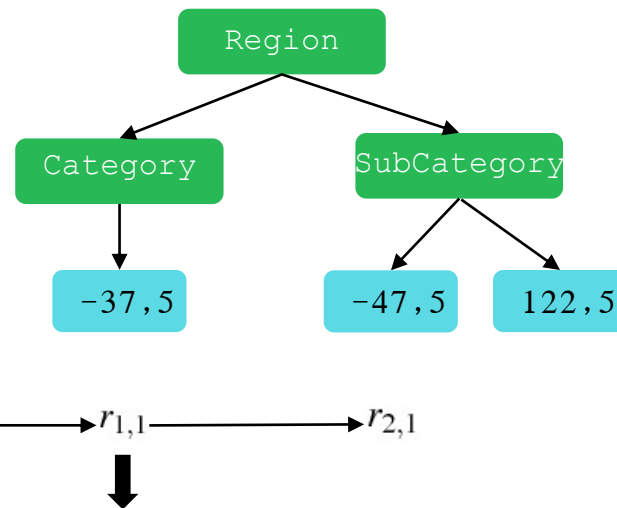


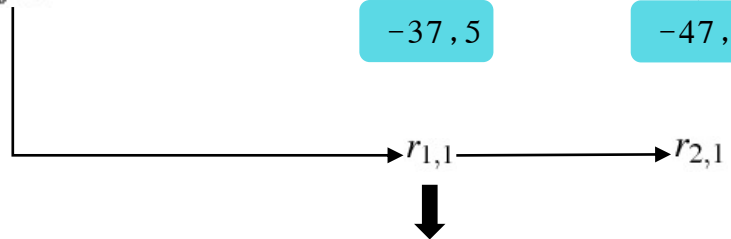
$$\gamma_{1,1} = \text{argmin} \frac{1}{2} (y_i - (F_{m-1}(x_i) + \gamma))^2 = \text{argmin} \frac{1}{2} (-37.5 - \gamma)^2 \quad \Rightarrow \quad \frac{\partial}{\partial \gamma} \frac{1}{2} (-37.5 - \gamma)^2$$

В цикле  $j = 1 \dots m$   
вычисление  $\gamma$   $\text{argmin}$

$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


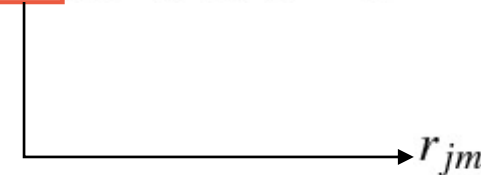
A diagram showing a line from the boxed  $\Sigma$  in the equation above, going down and then right to point at the variable  $r_{jm}$ .



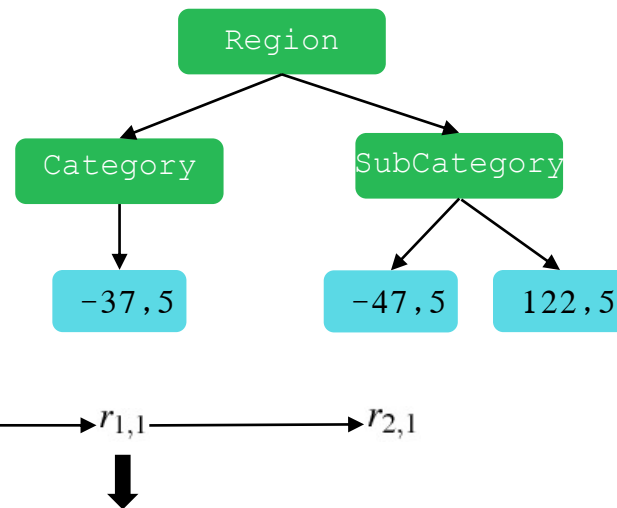
$$\frac{\partial}{\partial \gamma} \frac{1}{2} (-37.5 - \gamma)^2 \Rightarrow 37.5 + \gamma = 0$$


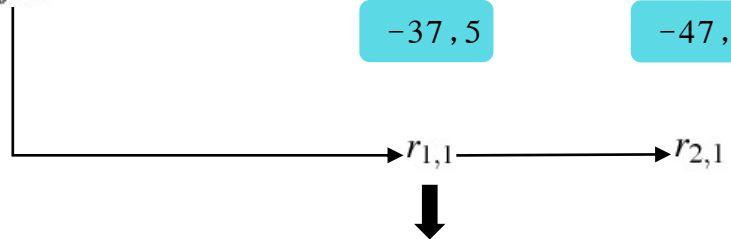
A diagram showing a line from  $r_{1,1}$  in the previous block, going down and then right to point at the derivative equation below.

В цикле  $j = 1 \dots m$   
вычисление  $\gamma$   $\text{argmin}$

$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


A diagram showing a line from the boxed  $\Sigma$  in the equation above, going down and then right to point at the variable  $r_{jm}$ .



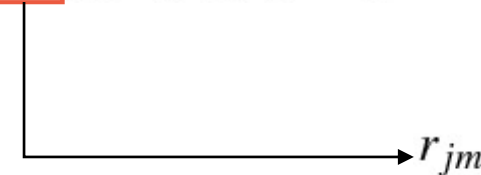
$$\frac{\partial}{\partial \gamma} \frac{1}{2} (-37.5 - \gamma)^2 \Rightarrow 37.5 + \gamma = 0 \Rightarrow -37.5$$


A diagram showing a line from the variable  $r_{1,1}$  in the block above, going down and then right to point at the derivative equation below.

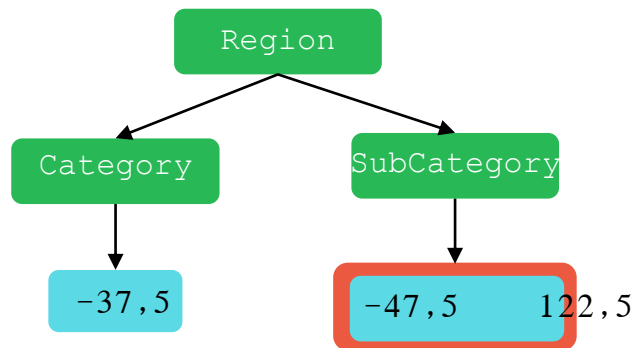
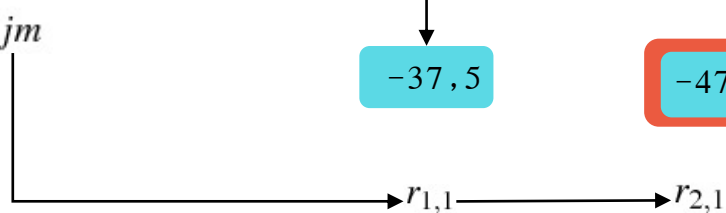
В цикле  $j = 1 \dots m$

**вычисление**  $\gamma$   $\text{argmin}$

(если несколько значений в листе)

$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


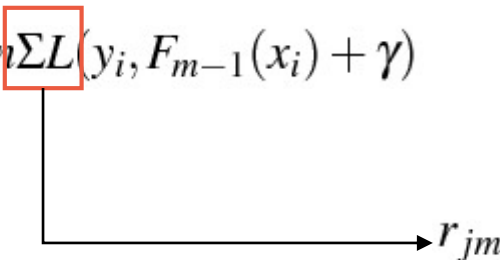
A diagram showing a line from the boxed  $\text{argmin}$  in the equation above, extending downwards and then to the right to point at the variable  $r_{jm}$ .



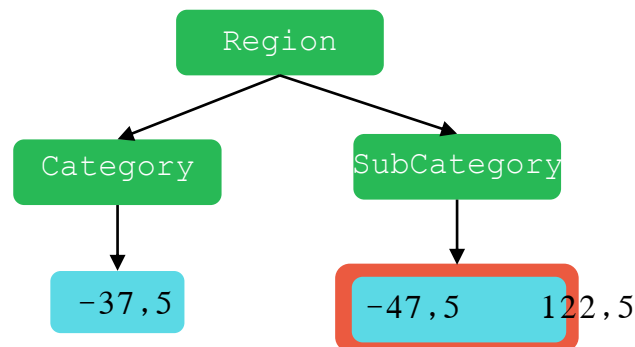
В цикле  $j = 1 \dots m$

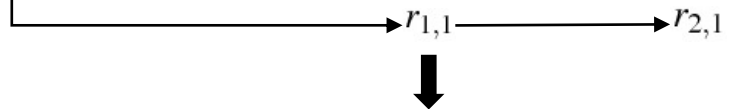
**вычисление**  $\gamma$   $\text{argmin}$

(если несколько значений в листе)

$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


The diagram shows a red box around the word "argmin" in the equation above. An arrow points from this box down and then right to the variable  $r_{jm}$ .



$$r_{1,1} \quad r_{2,1}$$


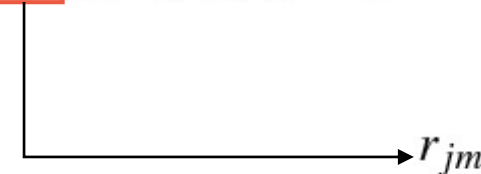
The diagram shows arrows from  $r_{1,1}$  and  $r_{2,1}$  pointing to the final calculation. A thick downward arrow is also present between  $r_{1,1}$  and the final calculation.

$$47.5 + \gamma - 122.5 + \gamma = 0 \rightarrow \frac{47.5 - 122.5}{2} = -37.5$$

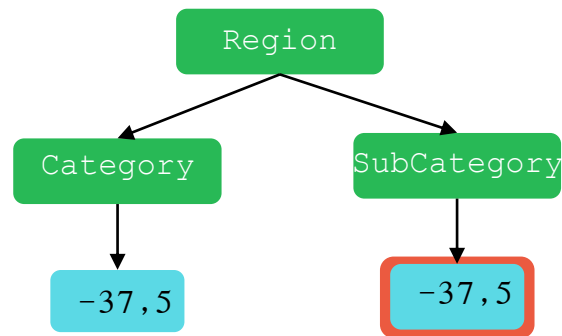
В цикле  $j = 1 \dots m$


**вычисление**  $\gamma$   $\text{argmin}$

(если несколько значений в листе)

$$\gamma_{jm} = \text{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$


The diagram shows a line starting from the boxed  $\Sigma$  in the equation above, going down and then right to point at the variable  $r_{jm}$ .



$$r_{1,1} \quad r_{2,1}$$


The diagram shows two arrows: one pointing down from  $r_{1,1}$  and one pointing up from  $r_{2,1}$ , both pointing towards the calculation below.

$$47.5 + \gamma - 122.5 + \gamma = 0 \Rightarrow \frac{47.5 - 122.5}{2} = -37.5$$



# Обновляем результат

$$F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$$

# Обновляем результат

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

На примере шага 1, где  $m = 1$ ,  
 $v = \text{Learning rate}$

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5$$

# Обновляем результат

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

На примере шага 1, где  $m = 1$ ,  $v =$   
Learning rate

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5 + \text{Learning Rate}$$

# Обновляем результат

$$F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$$

На примере шага 1, где  $m = 1$ ,  $v =$   
Learning rate

$$F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$$

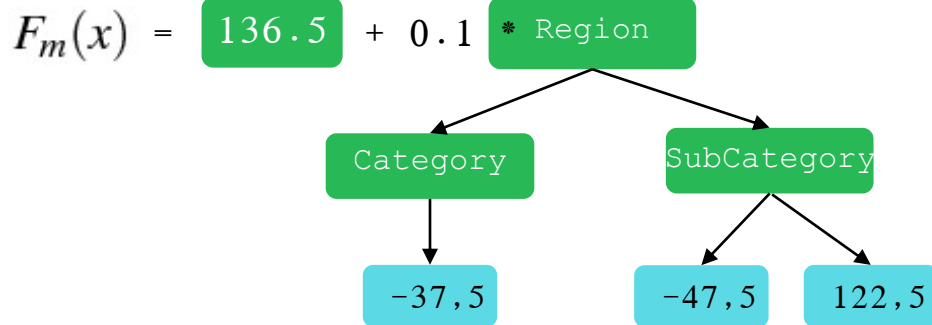
$$F_m(x) = 136.5 + 0.1$$

# Обновляем результат

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

На примере шага 1, где  $m = 1$ ,  $v =$   
Learning rate

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$



# Обновляем результат

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

На примере шага 1, где  $m = 1$ ,  $v =$   
Learning rate

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5 + 0.1 \cdot -37.5$$

# Обновляем результат

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

На примере шага 1, где  $m = 1$ ,  $v =$   
Learning rate

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5 + 0.1 \cdot -37.5 = 132.75$$

# Обновляем результат

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

На примере шага 1, где  $m = 1$ ,  $v =$   
Learning rate

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$

$$F_m(x) = 136.5 + 0.1 \cdot (-37.5) = 132.75$$



Region	Category	Sub category	Param1	threshold	price	diff	pred1
Владимирская область	Личные вещи	1	1	3	99	-37,5	132,75



Осталось  
сделать  
больше  
шагов



1  $Set(x_i, x_j)_{i=1}^n$  и функцию потерь  $L(y_i, F(x))$

2  $F_0(x) = argmin \Sigma L(y_i, \gamma)$

3 Вычисление  $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$

4 Обучение регрессии  $r_{jm}$  в цикле  $j = 1...m$

5 В цикле  $j = 1...m$  вычисление  $r_{jm} = argmin \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$

6  $F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$

**Посмотрим что  
мы посчитали**



# 3

## Градиент для классификации

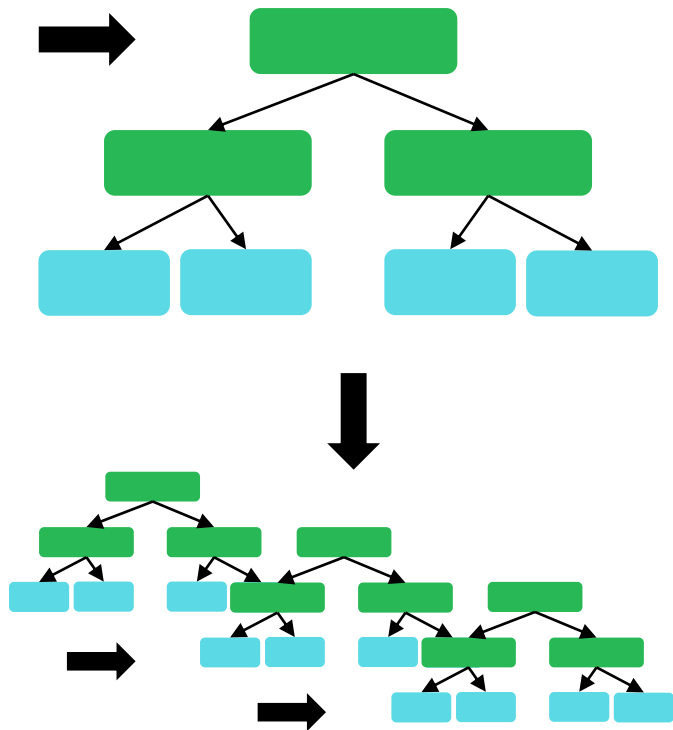
## В поиск класса

Region	Category	Sub category	Target
Владимирская область	Личные вещи	1	0
Волгоградская область	Личные вещи	2	0
Кировская область	Личные вещи	5	1

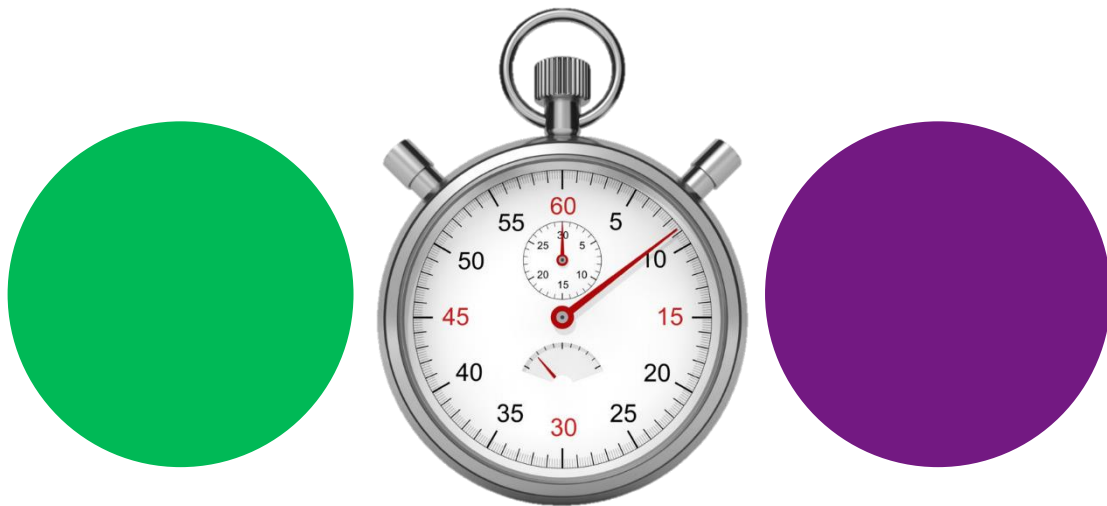


# В поиск класса

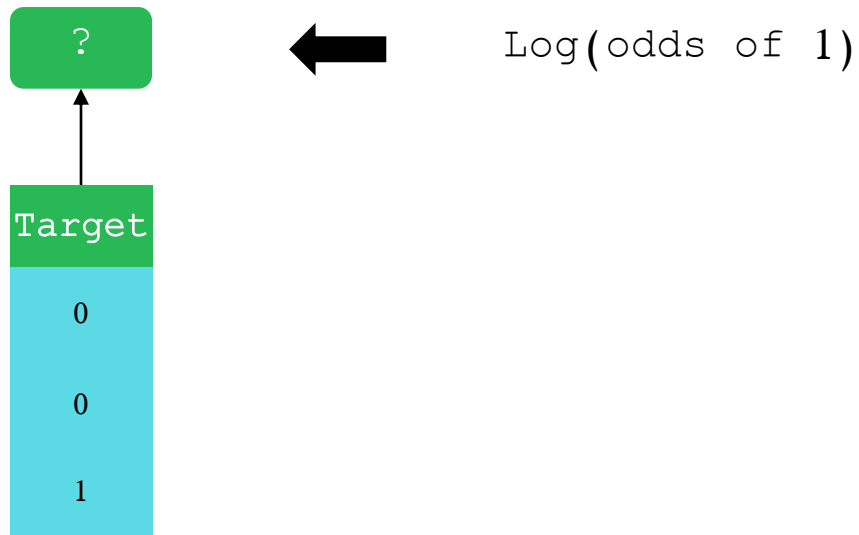
Region	Category	Sub category	Target
Владимирская область	Личные вещи	1	0
Волгоградская область	Личные вещи	2	0
Кировская область	Личные вещи	5	1



# Нужен старт для расчета



# Нужен старт для расчета



# Нужен старт для расчета

?



$\text{Log}(\text{count}(1) / \text{count}(0))$

Target

0

0

1





# Нужен старт для расчета

?



$\text{Log}(\text{count}(1) / \text{count}(0))$



$\log(\frac{1}{2})$

Target

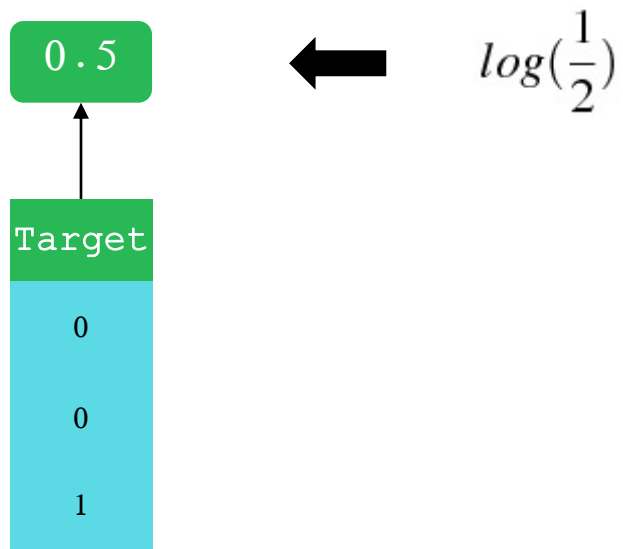
0

0

1



# Нужен старт для расчета



# Вероятность 1



$$\frac{e^{\log(\frac{|True|}{|False|})}}{1 + e^{\log(\frac{|True|}{|False|})}}$$

# Вероятность 1



$$\frac{e^{\log(\frac{|True|}{|False|})}}{1 + e^{\log(\frac{|True|}{|False|})}} = \frac{e^{\log(\frac{1}{2})}}{1 + e^{\log(\frac{1}{2})}} = 0.33$$

# Вероятность 1



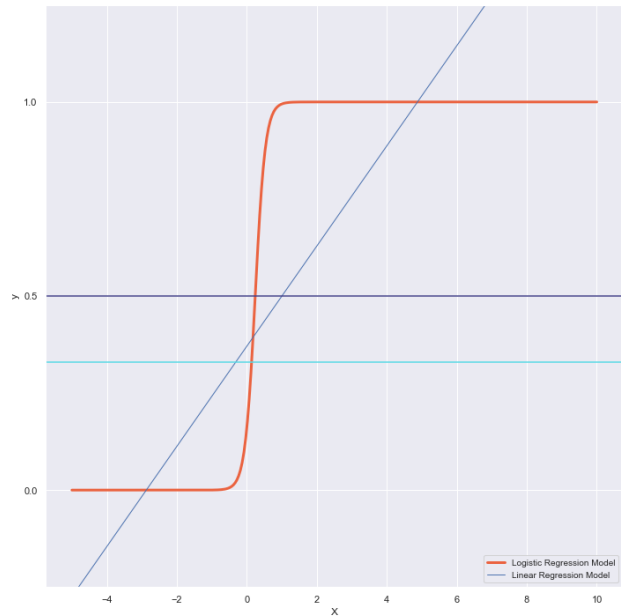
$$\frac{e^{\log(\frac{|True|}{|False|})}}{1 + e^{\log(\frac{|True|}{|False|})}} = 0.3 = \text{порог для } 1$$

# Вероятность 1

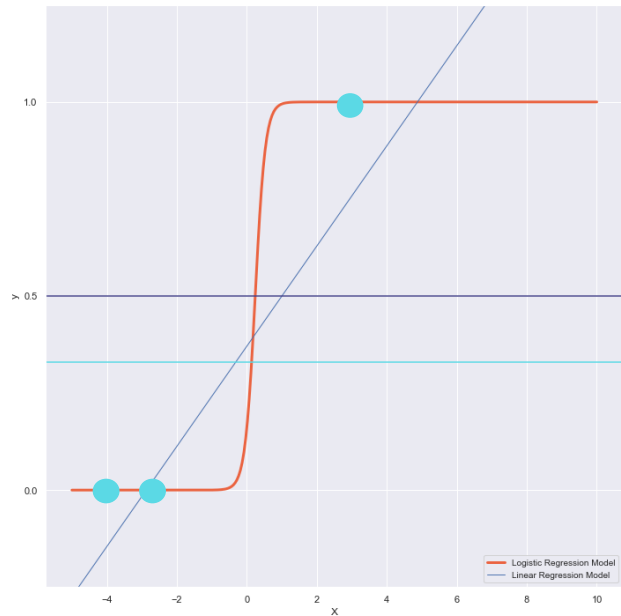
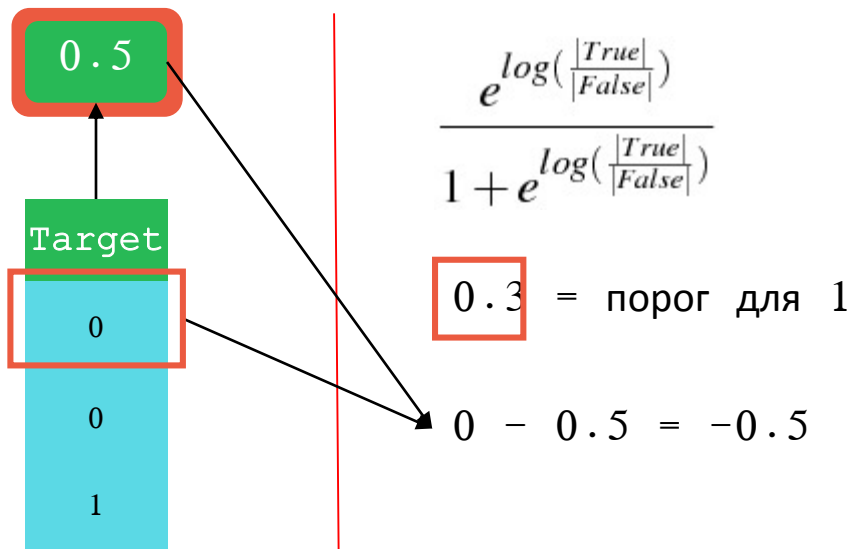


$$\frac{e^{\log\left(\frac{|True|}{|False|}\right)}}{1 + e^{\log\left(\frac{|True|}{|False|}\right)}}$$

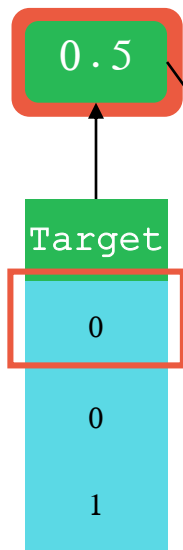
0.3 = порог для 1



# Вероятность 1



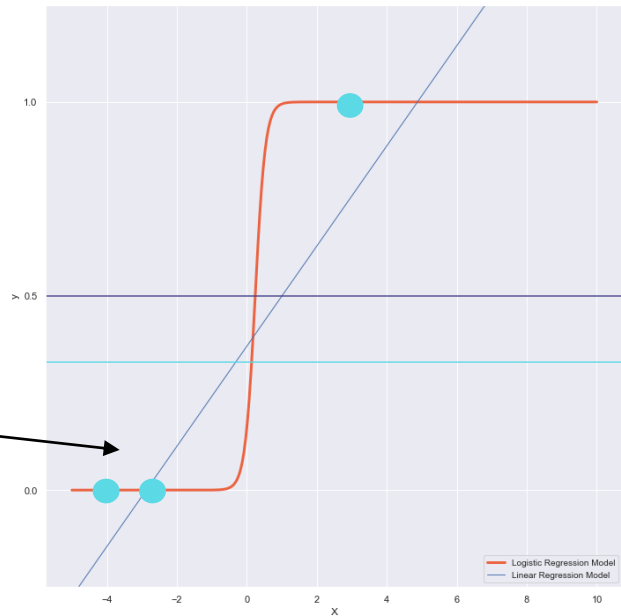
# Вероятность 1



$$\frac{e^{\log\left(\frac{|True|}{|False|}\right)}}{1 + e^{\log\left(\frac{|True|}{|False|}\right)}}$$

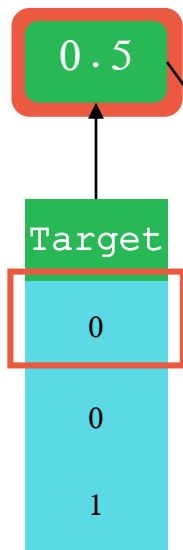
0.3 = порог для 1

$$0 - 0.5 = -0.5$$





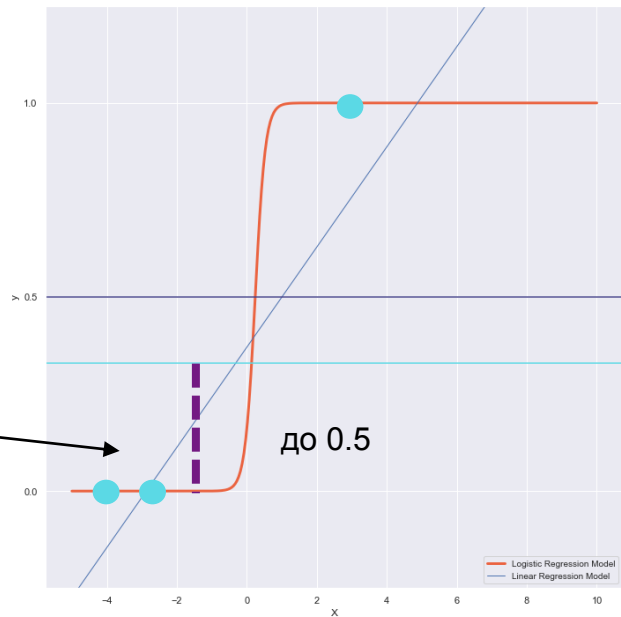
# Вероятность 1



$$\frac{e^{\log\left(\frac{|True|}{|False|}\right)}}{1 + e^{\log\left(\frac{|True|}{|False|}\right)}}$$

**0.3** = порог для 1

$$0 - 0.5 = -0.5$$



# Вероятность 1

Region	Category	Sub category	Target	diff
Владимирская область	Личные вещи	1	0	
Волгоградская область	Личные вещи	2	0	
Кировская область	Личные вещи	5	1	



(Observed - Predicted)

# Вероятность 1

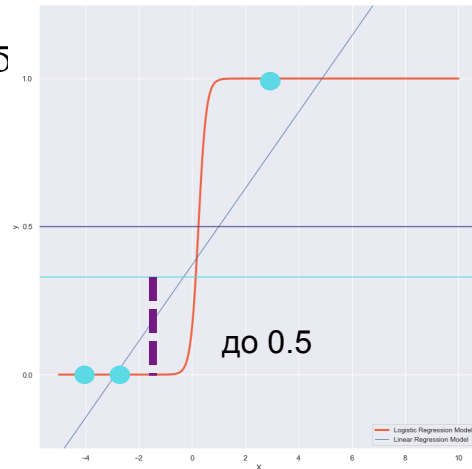
Region	Category	Sub category	Target	diff
Владимирская область	Личные вещи	1	0	-0,5
Волгоградская область	Личные вещи	2	0	-0,5
Кировская область	Личные вещи	5	1	0,5



$$(0 - 0.5)$$

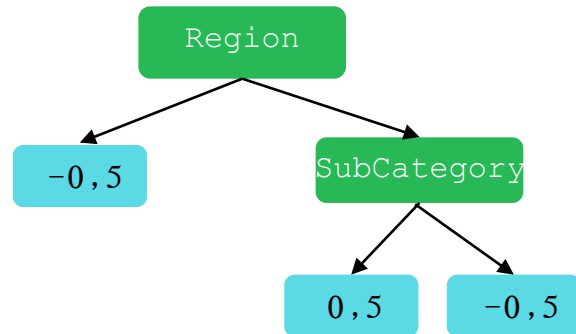
$$(0 - 0.5)$$

$$(1 - 0.5)$$



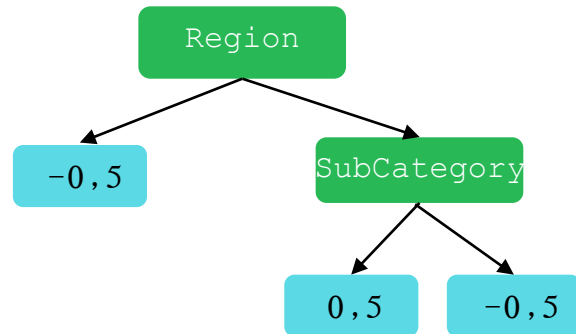
# Вероятность 1

Region	Category	Sub category	Target	diff
Владимирская область	Личные вещи	1	0	-0,5
Волгоградская область	Личные вещи	2	0	-0,5
Кировская область	Личные вещи	5	1	0,5



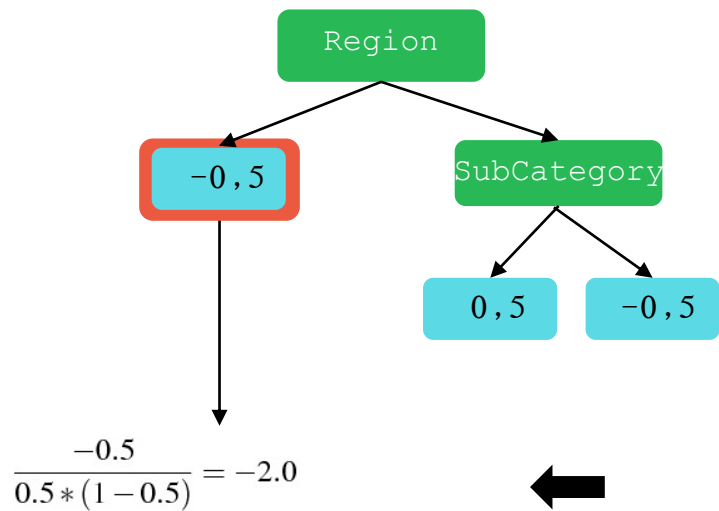
# Вероятность 1

Region	Category	Sub category	Target	diff
Владимирская область	Личные вещи	1	0	-0,5
Волгоградская область	Личные вещи	2	0	-0,5
Кировская область	Личные вещи	5	1	0,5



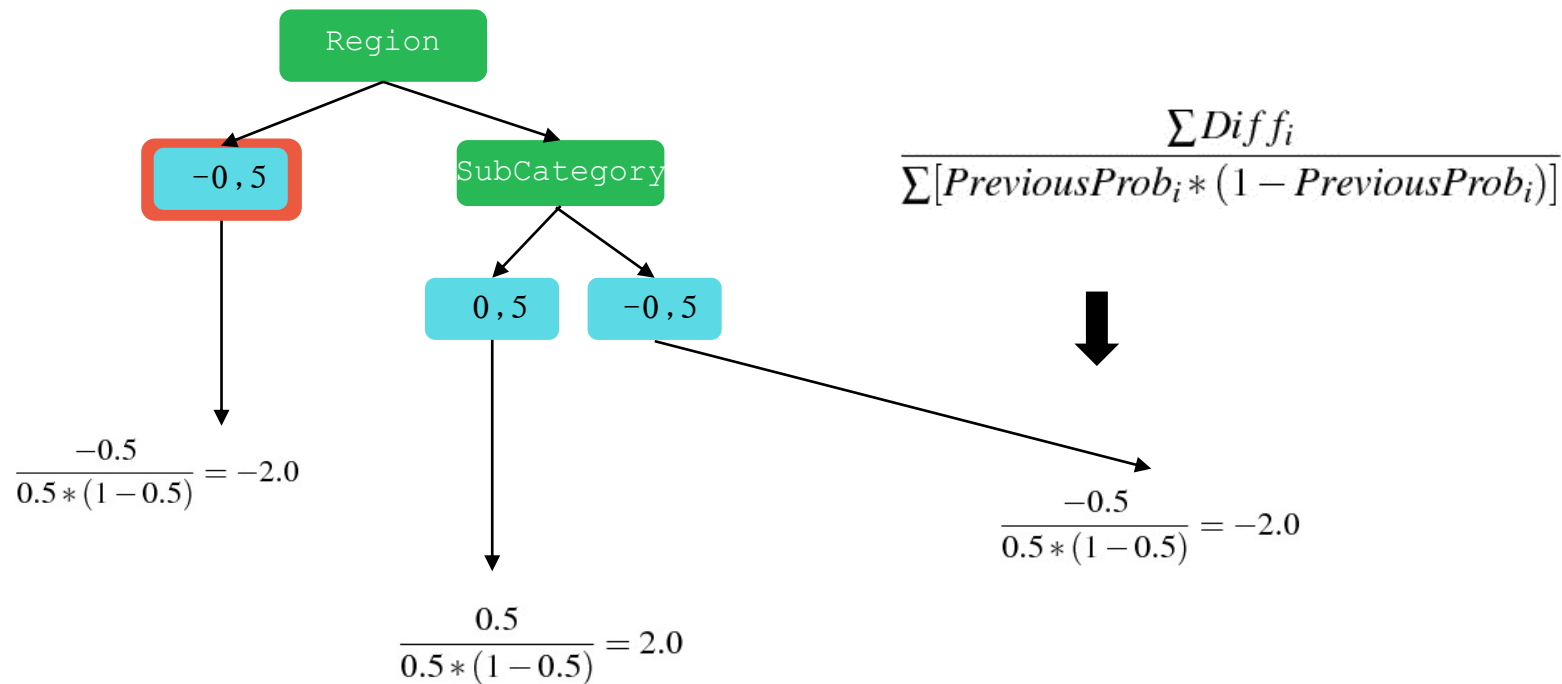
$$\frac{\sum Diff_i}{\sum [PreviousProb_i * (1 - PreviousProb_i)]}$$

# Вероятность

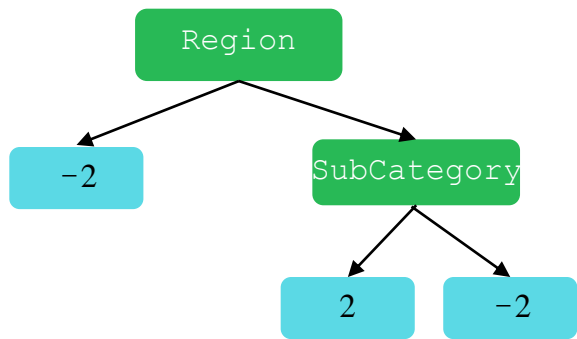


$$\frac{\sum Diff_i}{\sum [PreviousProb_i * (1 - PreviousProb_i)]}$$

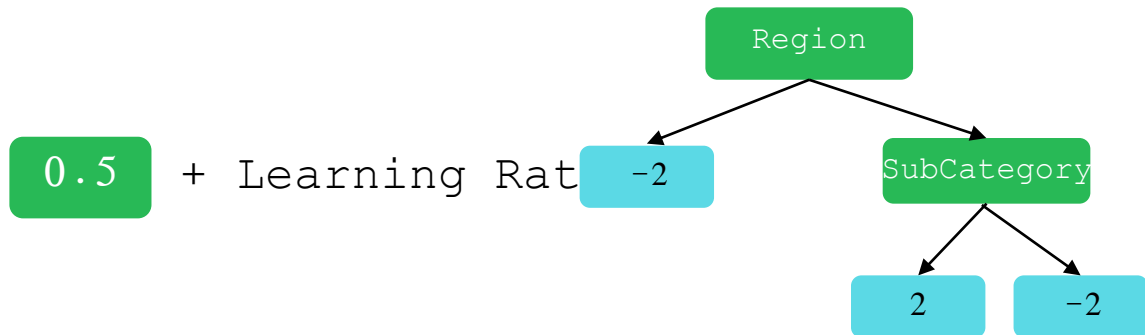
# Вероятность



# Вероятность



$$\frac{\sum Dif f_i}{\sum [PreviousProb_i * (1 - PreviousProb_i)]}$$



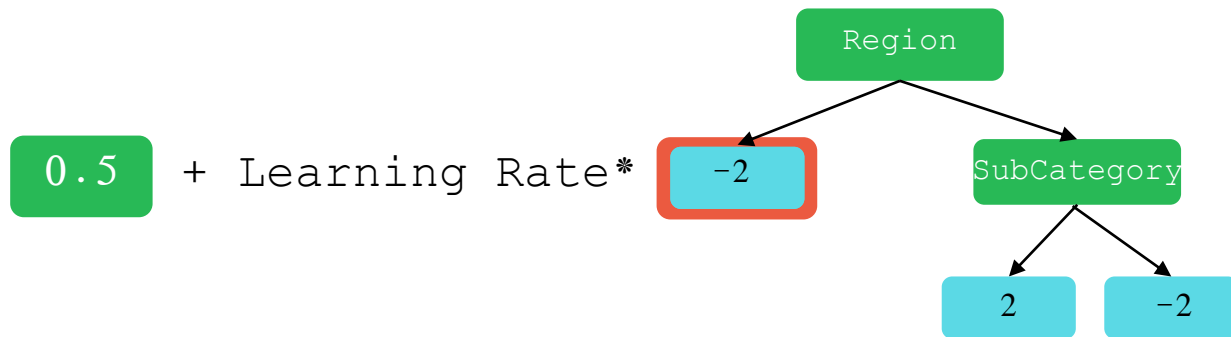


# Вероятность



$$\text{Predict} = 0.5 + (0.1 * -2) = 0.3$$

# Вероятность



$$\text{Predict} = 0.5 + (0.1 * -2) = 0.3$$





$$\frac{e^{\log(\frac{1}{2})}}{1 + e^{\log(\frac{1}{2})}} \rightarrow \frac{e^{0.3}}{1 + e^{0.3}} = 0.57$$

# Вероятность

$$\text{Predict} = 0.5 + (0.1 * \text{Diff})$$

Region	Category	Sub category	Target	Diff	Pred
Владимирская область	Личные вещи	1	0	-0,5	0,57
Волгоградская область	Личные вещи	2	0	-0,5	0,57
Кировская область	Личные вещи	5	1	0,5	0,66


$$\frac{e^{\text{New}}}{1 + e^{\text{New}}}$$


Predict Prob

# Вероятность

$$\text{Predict} = 0.5 + (0.1 * \text{Diff})$$

Region	Category	Sub category	Target	Diff	Pred
Владимирская область	Личные вещи	1	0	-0,5	0,57
Волгоградская область	Личные вещи	2	0	-0,5	0,57
Кировская область	Личные вещи	5	1	0,5	0,66

A diagram illustrating the conversion of a linear prediction into a probability. It starts with the equation  $\text{Predict} = 0.5 + (0.1 * \text{Diff})$  at the top. A large downward arrow points to the sigmoid function formula: 
$$\frac{e^{\text{New}}}{1 + e^{\text{New}}}$$
. From this formula, another large downward arrow points to the text "Predict Prob". A large leftward arrow points from "Predict Prob" to the word "Predict".

# Повторить много раз

Region	Category	Sub category	Target	pred
Владимирская область	Личные вещи	1	0	↓
Волгоградская область	Личные вещи	2	0	↓
Кировская область	Личные вещи	5	1	↑

# 4

**Больше  
деталей**

# Разобрали на примере



1  $Set(x_i, x_j)_{i=1}^n$  и функцию потерь  $L(y_i, F(x))$

2  $F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$

3 Вычисление  $r_{i,m} = -[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$

4 Обучение регрессии  $r_{jm}$  в цикле  $j = 1 \dots m$

5 В цикле  $j = 1 \dots m$  вычисление  $\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$

6  $F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$

## Что изменилось

$$F_0(x) = \boxed{\operatorname{argmin} \Sigma L(y_i, \gamma)}$$



для регрессии

$$\frac{1}{2}(\textit{Observed} - \textit{Predicted})^2 = \frac{2}{2}(\textit{Observed} - \textit{Predicted})x - 1 = -(\textit{Observed} - \textit{Predicted})$$



# Что изменилось

$$F_0(x) = \boxed{\operatorname{argmin} \sum L(y_i, \gamma)}$$



для регрессии

$$\frac{1}{2}(\text{Observed} - \text{Predicted})^2 = \frac{2}{2}(\text{Observed} - \text{Predicted})x - 1 = -(\text{Observed} - \text{Predicted})$$



для классификации

$\text{Log}(\text{likelihood of Observed Data})$

# Вероятность

$$F_0(x) = \boxed{\operatorname{argmin} \sum L(y_i, \gamma)}$$



для регрессии

$$\frac{1}{2}(\text{Observed} - \text{Predicted})^2 = \frac{2}{2}(\text{Observed} - \text{Predicted})x - 1 = -(\text{Observed} - \text{Predicted})$$



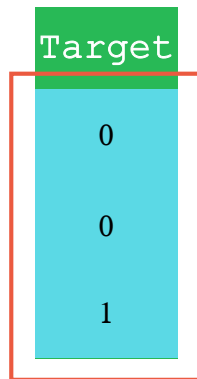
для классификации

$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$

p =

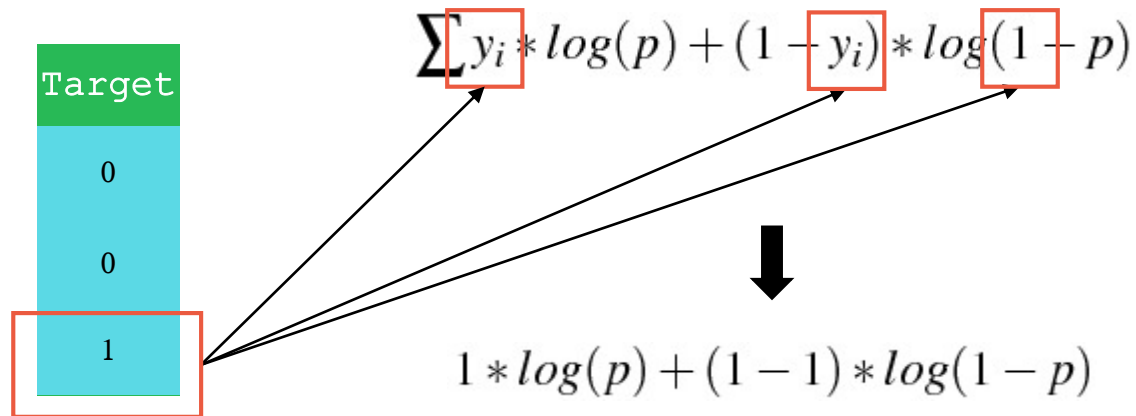
0.5

# Рассмотрим подробнее $\log(\text{likelihood})$

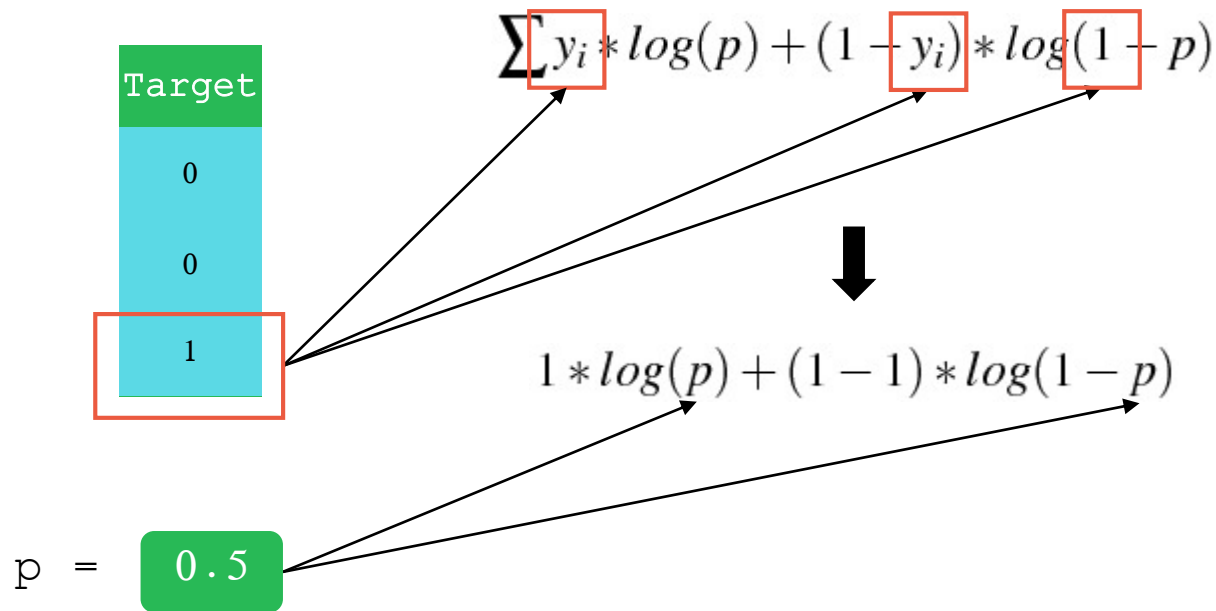


$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$

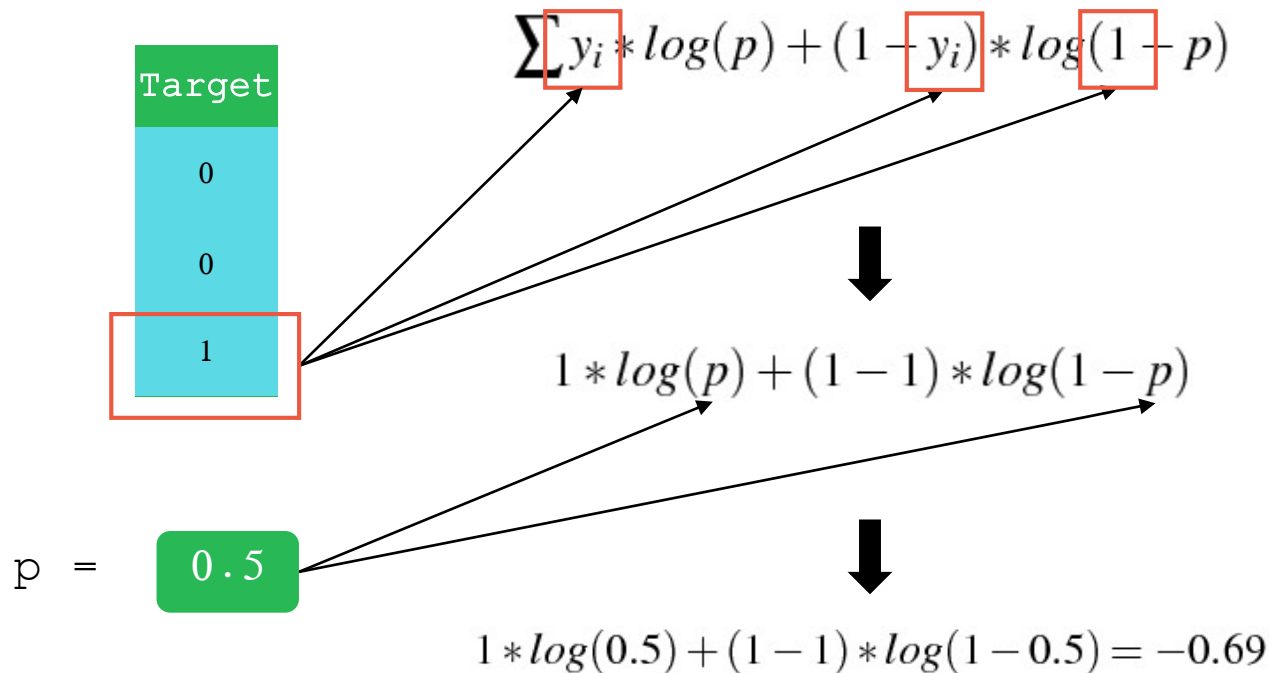
# Рассмотрим подробнее $\log(\text{likelihood})$



# Рассмотрим подробнее $\log(\text{likelihood})$

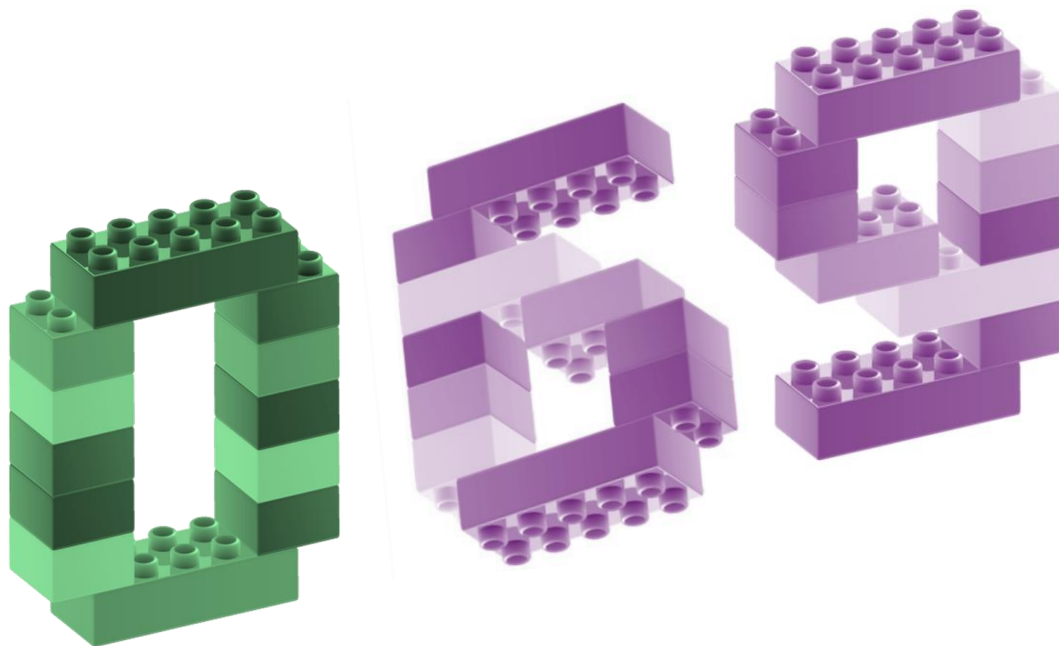


# Рассмотрим подробнее $\log(\text{likelihood})$



– 0.69?

$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$



# Упростим Loss

$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p) \quad \longrightarrow \quad -\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$





# Упростим Loss

$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p) \quad \rightarrow \quad -\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$



$$-\boxed{[Observed * \log(p) + (1 - Observed) * \log(1 - p)]}$$



# Упростим Loss

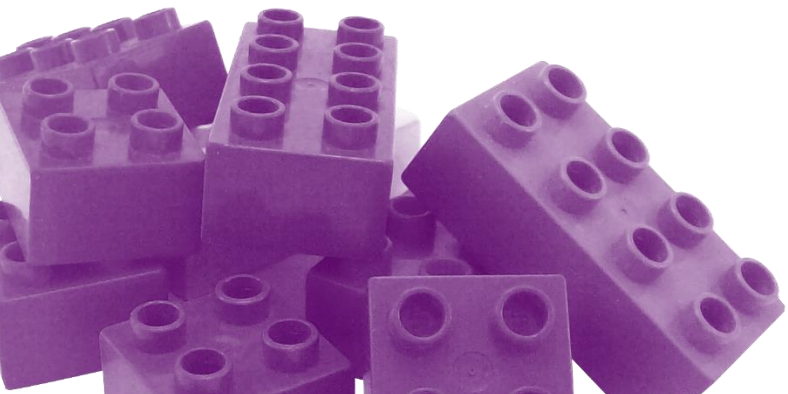
$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p) \quad \rightarrow \quad -\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$



$$-[\text{Observed} * \log(p) + (1 - \text{Observed}) * \log(1 - p)]$$



$$-\text{Observed} * \log(p) - (1 - \text{Observed}) * \log(1 - p)$$



# Упростим Loss

$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p) \quad \rightarrow \quad -\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$



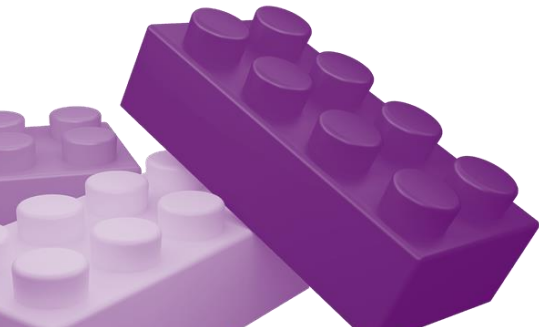
$$-[\text{Observed} * \log(p) + (1 - \text{Observed}) * \log(1 - p)]$$



$$-\text{Observed} * \log(p) - (1 - \text{Observed}) * \log(1 - p)$$



$$-\text{Observed} * \log(p) - \log(1 - p) + \text{Observed} * \log(1 - p)$$



# Упростим Loss

$$\sum y_i * \log(p) + (1 - y_i) * \log(1 - p) \quad \Rightarrow \quad -\sum y_i * \log(p) + (1 - y_i) * \log(1 - p)$$



$$-[\text{Observed} * \log(p) + (1 - \text{Observed}) * \log(1 - p)]$$



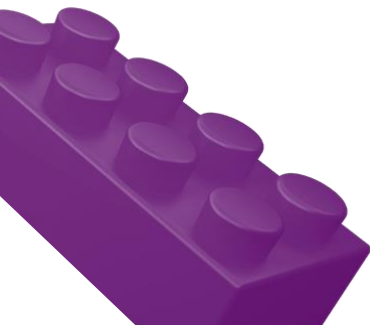
$$-\text{Observed} * \log(p) - (1 - \text{Observed}) * \log(1 - p)$$



$$-\text{Observed} * \log(p) - \log(1 - p) + \text{Observed} * \log(1 - p)$$



$$-\text{Observed} * [\log(p) - \log(1 - p)] - \log(1 - p)$$



## Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



## Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



$$\log(p) - \log(1 - p) = \frac{\log(p)}{\log(1 - p)}$$



## Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



$$\log(p) - \log(1 - p) = \frac{\log(p)}{\log(1 - p)}$$



$$\log\left(\frac{p}{1 - p}\right) = \log(\text{odds of } 1)$$



# Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



$$\log(p) - \log(1 - p) = \frac{\log(p)}{\log(1 - p)}$$



$$\log\left(\frac{p}{1 - p}\right) = \log(\text{odds of } 1)$$



$$\log(1 - p) = \log\left(1 - \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}\right)$$





# Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



$$\log(p) - \log(1 - p) = \frac{\log(p)}{\log(1 - p)}$$



$$\log\left(\frac{p}{1 - p}\right) = \log(\text{odds of } 1)$$



$$\log(1 - p) = \log\left(1 - \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}\right)$$



# Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



$$\log(p) - \log(1 - p) = \frac{\log(p)}{\log(1 - p)}$$



$$\log\left(\frac{p}{1 - p}\right) = \log(\text{odds of } 1)$$



$$\log(1 - p) = \log\left(1 - \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}\right)$$



$$\log\left(\frac{1 + e^{\log(odds)}}{1 + e^{\log(odds)}} - \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}\right)$$



# Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



$$\log(p) - \log(1 - p) = \frac{\log(p)}{\log(1 - p)}$$



$$\log\left(\frac{p}{1 - p}\right) = \log(\text{odds of } 1)$$



$$\log(1 - p) = \log\left(1 - \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}\right)$$



$$\log\left(\frac{1 + e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}} - \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}\right)$$



$$\log\left(\frac{1}{1 + e^{\log(\text{odds})}}\right) = \log(1) - \log(1 + e^{\log(\text{odds})})$$



# Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



$$\log(p) - \log(1 - p) = \frac{\log(p)}{\log(1 - p)}$$

$$\log\left(\frac{p}{1 - p}\right) = \log(\text{odds of } 1)$$

$$\log(1 - p) = \log\left(1 - \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}\right)$$

$$\log\left(\frac{1 + e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}} - \frac{e^{\log(\text{odds})}}{1 + e^{\log(\text{odds})}}\right)$$

$$\log\left(\frac{1}{1 + e^{\log(\text{odds})}}\right) = \log(1) - \log(1 + e^{\log(\text{odds})})$$

$$-\log(1 + e^{\log(\text{odds})})$$

## Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



$$\log\left(\frac{p}{1-p}\right) = \log(\text{odds of } 1)$$



$$-\log(1 + e^{\log(\text{odds})})$$



## Ещё несколько шагов

$$-Observed * [\log(p) - \log(1 - p)] - \log(1 - p)$$



$$\log\left(\frac{p}{1-p}\right) = \log(\text{odds of } 1)$$

$$-\log(1 + e^{\log(odds)})$$



$$-Observed * \log(odds) + \log(1 + e^{\log(odds)})$$



Возьмем производную (от Loss

$$\frac{\partial}{\partial \log(odds)} - Observed * \log(odds) + \log(1 + e^{\log(odds)})$$

# Возьмем производную от Loss

$$\frac{\partial}{\partial \log(odds)} - Observed * \log(odds) + \log(1 + e^{\log(odds)})$$



- Observed



## Возьмем производную от Loss

$$\frac{\partial}{\partial \log(odds)} - Observed * \log(odds) + \log(1 + e^{\log(odds)})$$



- Observed



$$\frac{1}{1 + e^{\log(odds)}}$$

Возьмем производную от LOSS

$$\frac{\partial}{\partial \log(odds)} - Observed * \log(odds) + \log(1 + e^{\log(odds)})$$



$$-Observed + \frac{1}{1 + e^{\log(odds)}} * e^{\log(odds)}$$

Возьмем производную от LOSS

$$\frac{\partial}{\partial \log(odds)} - Observed * \log(odds) + \log(1 + e^{\log(odds)})$$



$$-Observed + \frac{1}{1 + e^{\log(odds)}} * e^{\log(odds)}$$

## Возьмем производную от Loss

$$\frac{\partial}{\partial \log(odds)} - Observed * \log(odds) + \log(1 + e^{\log(odds)})$$



$$-Observed + \frac{1}{1 + e^{\log(odds)}} * e^{\log(odds)}$$



$$-Observed + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$



**Градиентный  
спуск для  
классификации**

Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$

price
99
99
89
259



$$-0 * \log(odds) + \log(1 + e^{\log(odds)})$$



$$-0 * \log(odds) + \log(1 + e^{\log(odds)})$$



$$-0 * \log(odds) + \log(1 + e^{\log(odds)})$$



Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$

price
99
99
89
259



$$-0 * \log(odds) + \log(1 + e^{\log(odds)}) \longrightarrow -0 + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$



$$-0 * \log(odds) + \log(1 + e^{\log(odds)}) \longrightarrow -0 + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$



$$-0 * \log(odds) + \log(1 + e^{\log(odds)}) \longrightarrow -0 + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$

Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$

price
99
99
89
259



$$-0 * \log(odds) + \log(1 + e^{\log(odds)}) \longrightarrow -0 + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}} \longrightarrow -0 + p$$



$$-0 * \log(odds) + \log(1 + e^{\log(odds)}) \longrightarrow -0 + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}} \longrightarrow -0 + p$$



$$-0 * \log(odds) + \log(1 + e^{\log(odds)}) \longrightarrow -0 + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}} \longrightarrow -1 + p$$



Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$



$$-0 + p + (-0) + p + (-1) + p = 0$$

price
99
99
89
259



Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$

price
99
99
89
259



$$-0 + p + (-0) + p + (-1) + p = 0$$



$$-1 + 3p$$

$$3p = 1$$

$$p = 1/3 = 0.33$$



Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$

price
99
99
89
259



$$-0 + p + (-0) + p + (-1) + p = 0$$



$$p = \frac{1}{3} \Rightarrow \log\left(\frac{p}{1-p}\right)$$



# Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$

price
99
99
89
259



$$-0 + p + (-0) + p + (-1) + p = 0$$



$$p = \frac{1}{3} \Rightarrow \log\left(\frac{p}{1-p}\right)$$



$$\log\left(\frac{\frac{1}{3}}{1-\frac{1}{3}}\right) \Rightarrow -0.69$$



Argmin <- минимизируем потерю

$$F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$$



$$F_0(x) = -0.69$$

price
99
99
89
259



# Теперь разберём



1  $Set(x_i, x_j)_{i=1}^n$  и функцию потерь  $L(y_i, F(x))$

2  $F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$

3 Вычисление

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x) = F_{m-1}(x)}$$

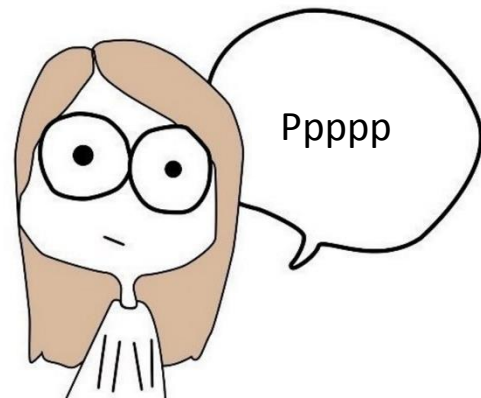
4 Обучение регрессии  $r_{jm}$  в цикле  $j = 1 \dots m$

5 В цикле  $j = 1 \dots m$  вычисление  $r_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$

6  $F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$

# Вычислим $\mathbb{R}$

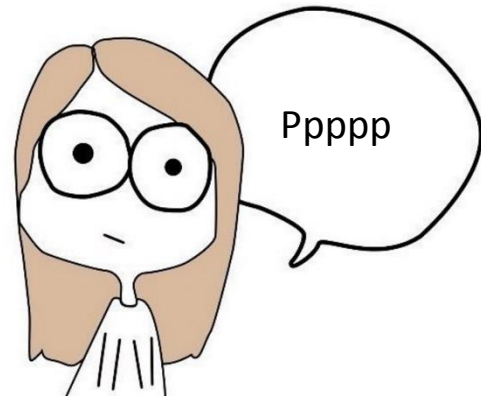
$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$



# Вычислим R

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

$$-Observed + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$





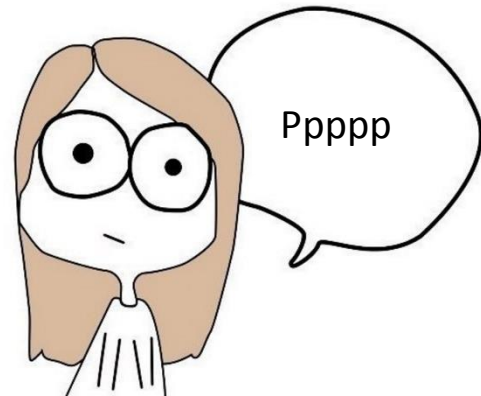
# Вычислим R

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}$$

↓

$$-Observed + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$

→ - 0.69



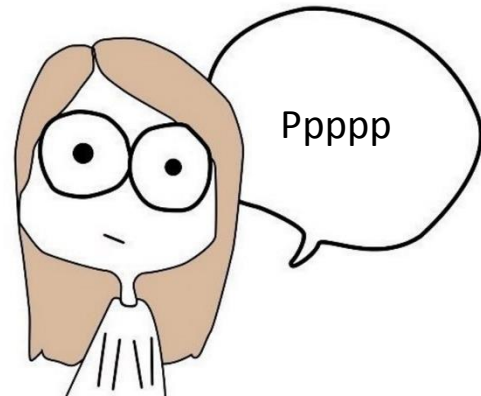
# Вычислим R

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} - 0.69$$

$$-Observed + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$



$$Observed - \frac{e^{-0.69}}{1 + e^{-0.69}}$$



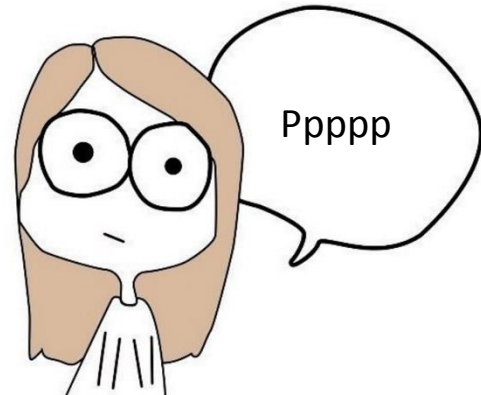
# Вычислим R

$$r_{i,m} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)} - 0.69$$

$$-Observed + \frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$$



$$Observed - \frac{e^{-0.69}}{1 + e^{-0.69}} \longrightarrow Observed - p = Observed - 0.33$$



# Вычислим R

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

Region	Category	Sub category	Target	R
Владимирская область	Личные вещи	1	0	-0,33
Волгоградская область	Личные вещи	2	0	-0,33
Кировская область	Личные вещи	5	1	0,67

← Observed - 0.33

← Observed - 0.33

← Observed - 0.33

# Вычислим R

$$r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$$

Region	Category	Sub category	Target	R
Владимирская область	Личные вещи	1	0	-0,33
Волгоградская область	Личные вещи	2	0	-0,33
Кировская область	Личные вещи	5	1	0,67

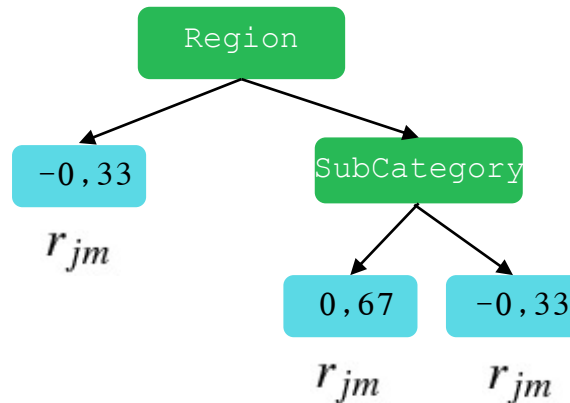
← 0 - 0.33

← 0 - 0.33

← 1 - 0.33

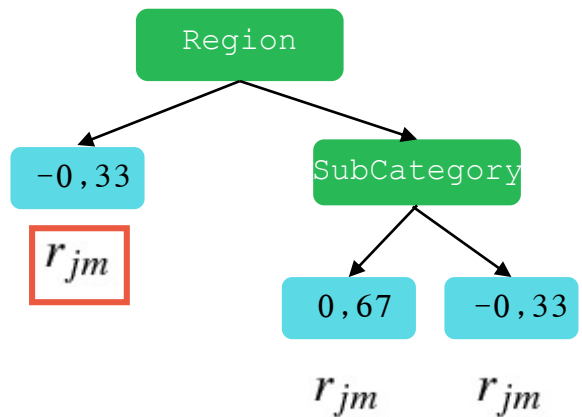
# Обучение регрессии на дереве

Region	Category	Sub category	Target	R
Владимирская область	Личные вещи	1	0	-0,33
Волгоградская область	Личные вещи	2	0	-0,33
Кировская область	Личные вещи	5	1	0,67



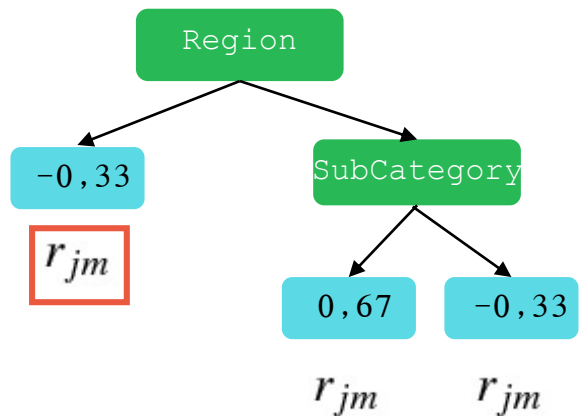
# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

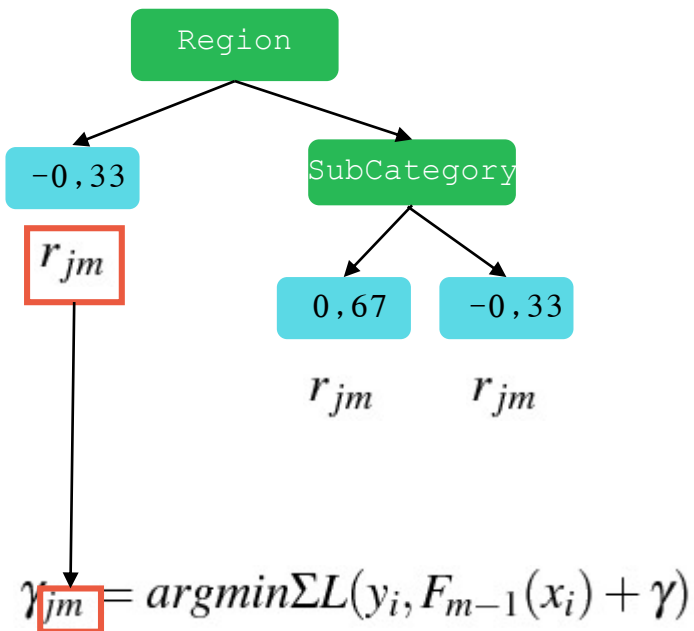


$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



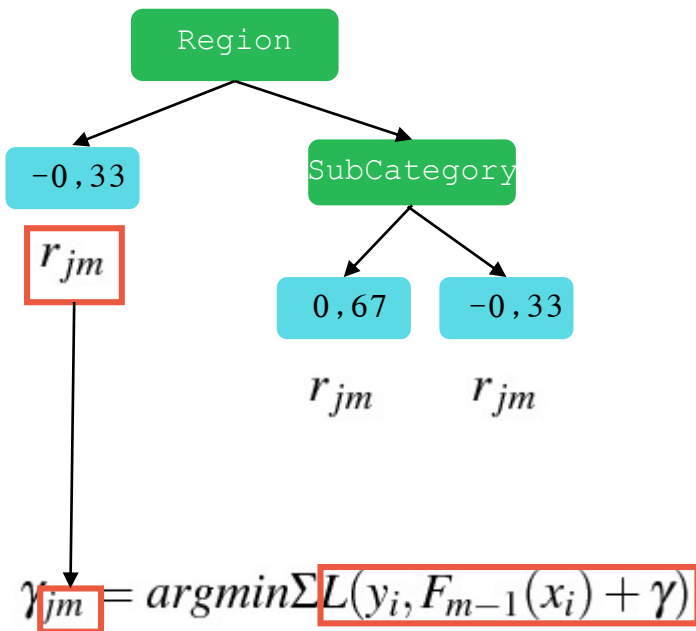
# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



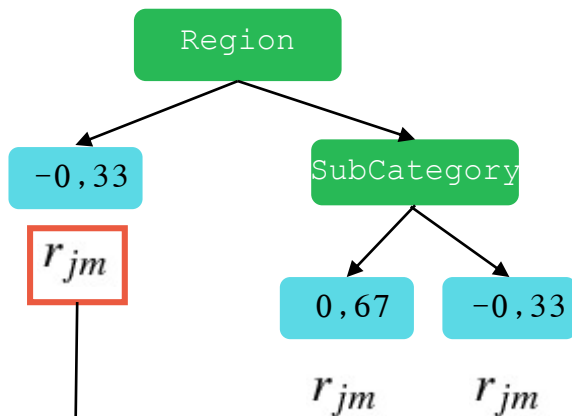
# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



# Вычисление

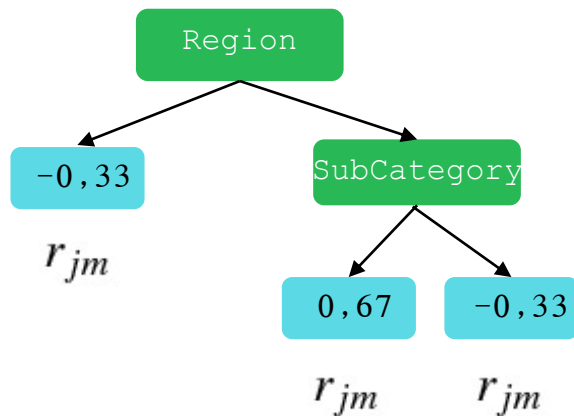
$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma) \rightarrow -\text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



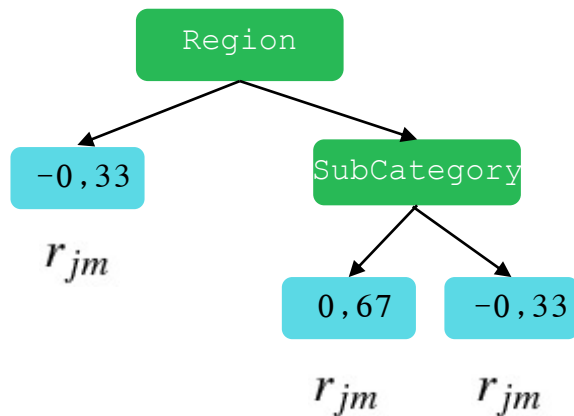
$$\operatorname{argmin} \Sigma -y_i * [F_{m-1}(x_i) + \gamma] + \log(1 + e^{F_{m-1}(x_i) + \gamma})$$

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

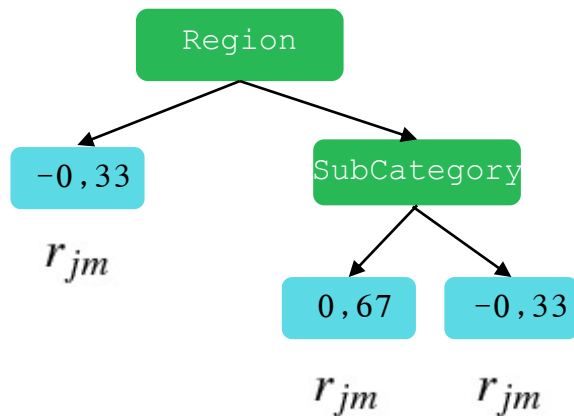
$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

$$\operatorname{argmin} \Sigma - \boxed{y_i} * [F_{m-1}(x_i) + \gamma] + \log(1 + e^{F_{m-1}(x_i) + \gamma})$$

-0,33

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



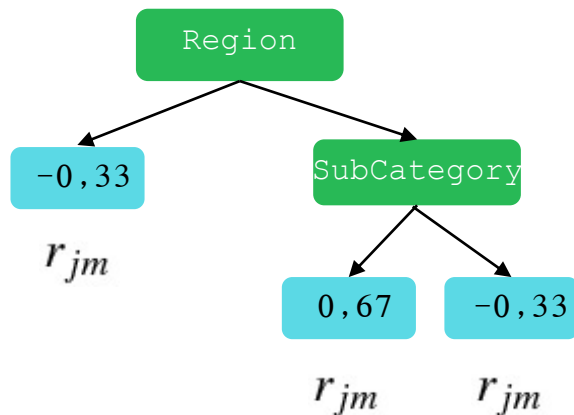
$$\operatorname{argmin} \Sigma -y_i * [F_{m-1}(x_i) + \gamma] + \log(1 + e^{F_{m-1}(x_i) + \gamma})$$

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

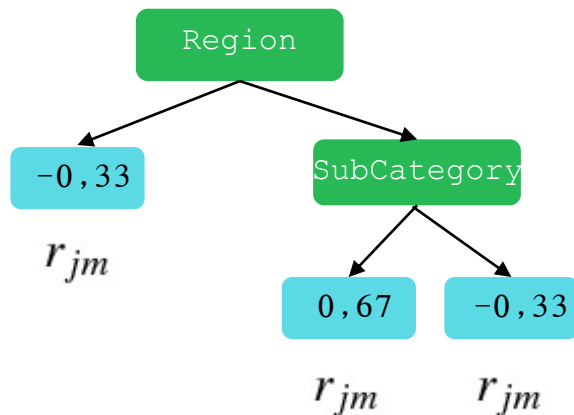
$$\operatorname{argmin} \Sigma - y_i * [F_{m-1}(x_i) + \gamma] + \log(1 + e^{F_{m-1}(x_i) + \gamma})$$



$$\frac{\partial}{\partial \gamma} L(y_i, F_{m-1}(x_i) + \gamma) \approx \frac{\partial}{\partial F()} (y_i, F_{m-1}(x_i)) + \frac{\partial^2}{\partial F()^2} (y_i, F_{m-1}(x_i)) \gamma$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

$$\operatorname{argmin} \Sigma - y_i * [F_{m-1}(x_i) + \gamma] + \log(1 + e^{F_{m-1}(x_i) + \gamma})$$



$$\frac{\partial}{\partial \gamma} L(y_i, F_{m-1}(x_i) + \gamma) \approx \frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i)) + \frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i)) \gamma$$

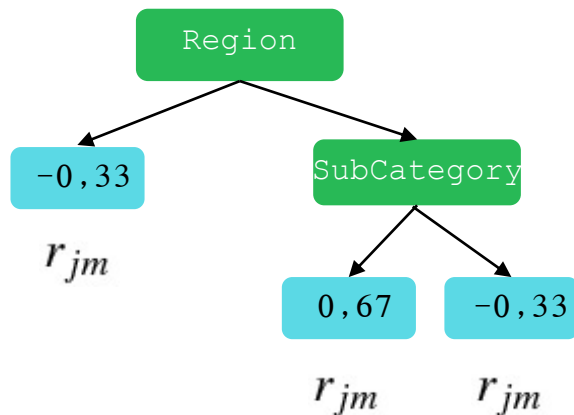


$$\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i)) \gamma = \frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))$$



# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

$$\operatorname{argmin} \Sigma - y_i * [F_{m-1}(x_i) + \gamma] + \log(1 + e^{F_{m-1}(x_i) + \gamma})$$



$$\frac{\partial}{\partial \gamma} L(y_i, F_{m-1}(x_i) + \gamma) \approx \frac{\partial}{\partial F()} (y_i, F_{m-1}(x_i)) + \frac{\partial^2}{\partial F()^2} (y_i, F_{m-1}(x_i)) \gamma$$



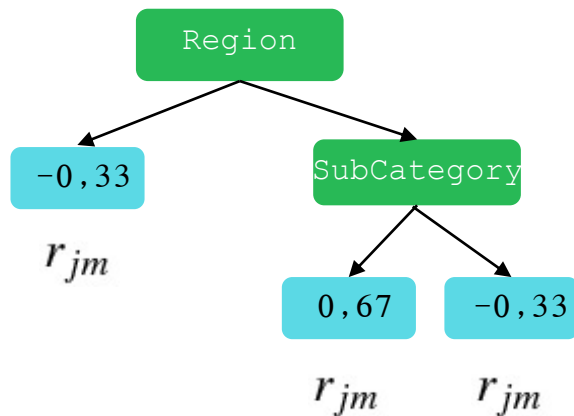
$$\frac{\partial^2}{\partial F()^2} (y_i, F_{m-1}(x_i)) \gamma = \frac{\partial}{\partial F()} (y_i, F_{m-1}(x_i))$$



$$\gamma = \frac{-\frac{\partial}{\partial F()} (y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2} (y_i, F_{m-1}(x_i))}$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



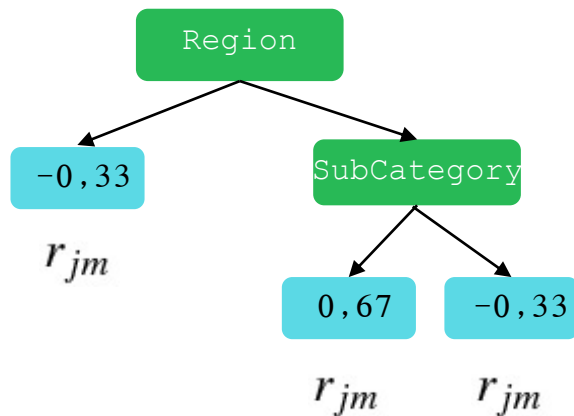
$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$-Observed * \log(odds) + \log(1 + e^{\log(odds)})$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

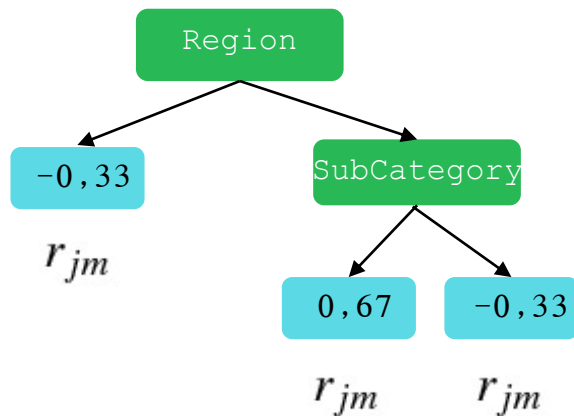
Observed - p



$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

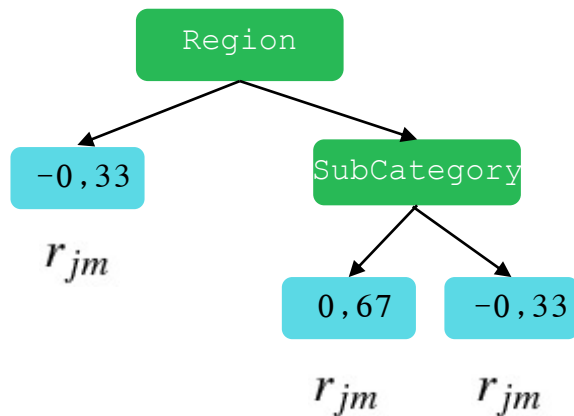
$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$



$$\frac{\partial^2}{\partial \log(\text{odds})^2} - \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

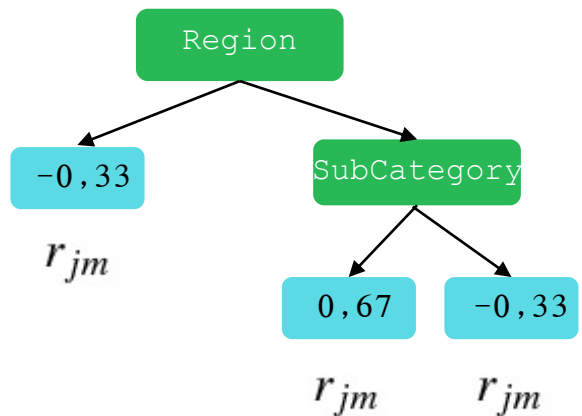
$$\gamma = \frac{-\frac{\partial}{\partial F()}(y_i, F_{m-1}(x_i))}{\frac{\partial^2}{\partial F()^2}(y_i, F_{m-1}(x_i))}$$



$$p * (1 - p)$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



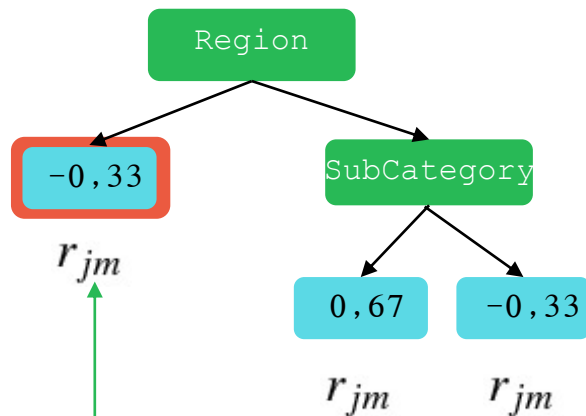
$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

$$\frac{\text{Observed} - p}{p * (1 - p)}$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



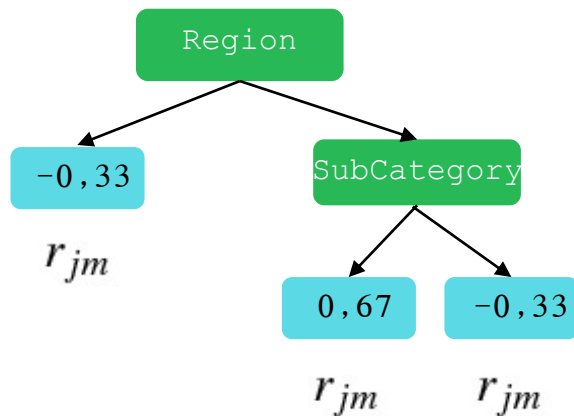
$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

$$\gamma_{1,1} = \frac{-0.33}{-0.69 * (1 - (-0.69))}$$

# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$

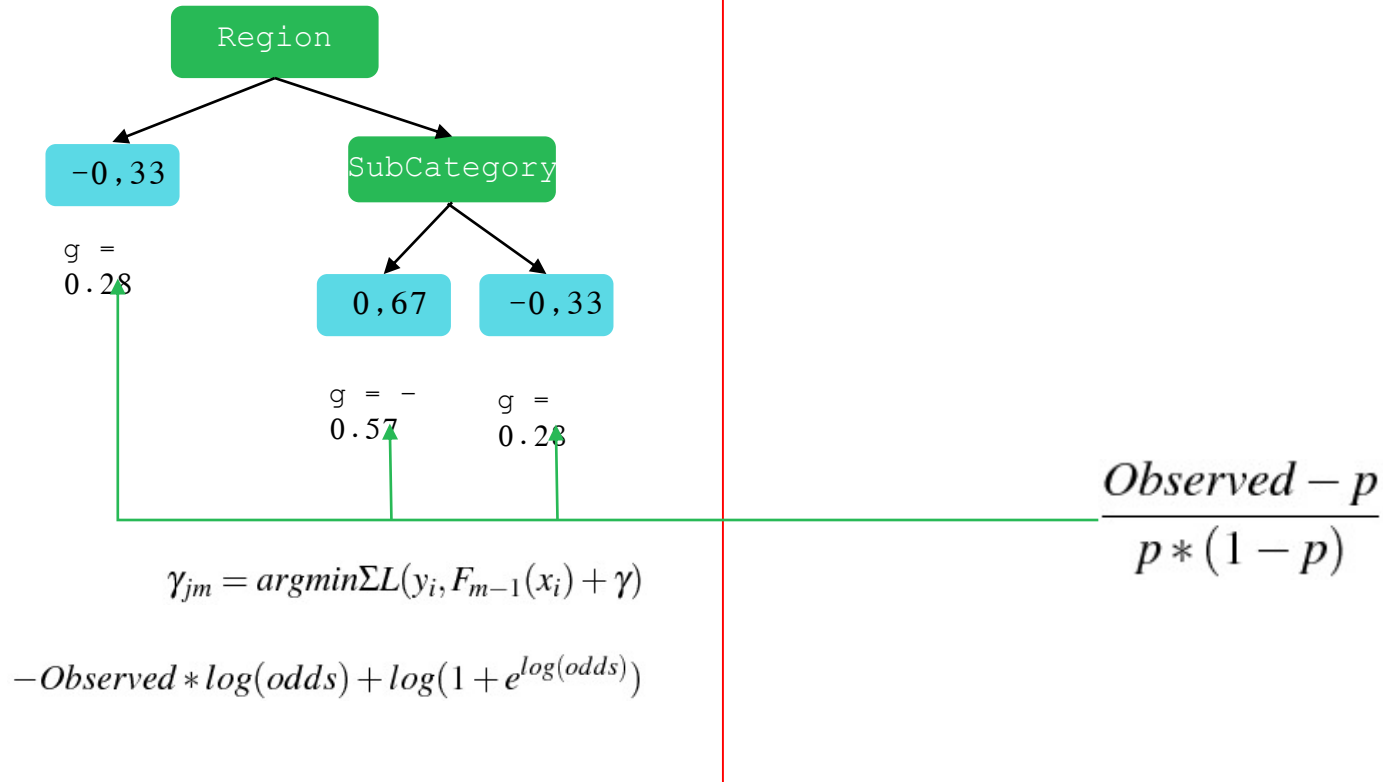
$$- \text{Observed} * \log(\text{odds}) + \log(1 + e^{\log(\text{odds})})$$

$$\gamma_{1,1} = \frac{-0.33}{-0.69 * (1 - (-0.69))} = 0.28$$



# Вычисление

$$\gamma_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$$



# Gamma для всех

$$\frac{Observed - p}{p * (1 - p)}$$



Region	Category	Sub category	Target	R	G
Владимирская область	Личные вещи	1	0	0,33	
Волгоградская область	Личные вещи	2	0	0,33	
Кировская область	Личные вещи	5	1	0,67	

← 0.28

← 0.28

← -0.57

# Обновляем $F$

$$F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$$

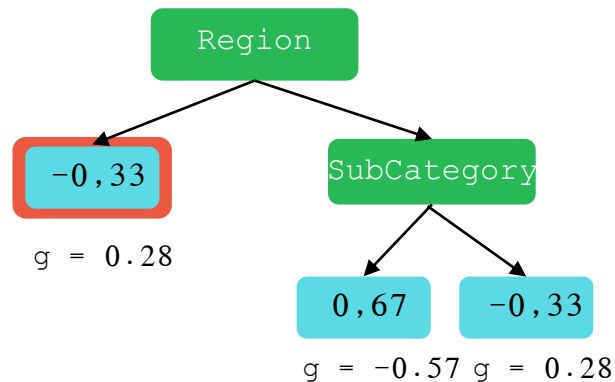
# Обновляем F

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$



# Обновляем $F$

$$F_m(x) = F_{m-1}(x) + v \sum \gamma_m I(x \in R_{jm})$$



$$F = 0.5 + (0.1 * 0.28) = 0.58$$

# Новый предикт для всех

NEW

Region	Category	Sub category	Target	Predict
Владимирская область	Личные вещи	1	0	0,58
Волгоградская область	Личные вещи	2	0	0,58
Кировская область	Личные вещи	5	1	0,44

←  $0.5 + (0.1 * 0.28) = 0.58$

←  $0.5 + (0.1 * 0.28) = 0.58$

←  $0.5 + (0.1 * -0.57) = 0.44$

# Завершили цикл



1  $Set(x_i, x_j)_{i=1}^n$  и функцию потерь  $L(y_i, F(x))$

2  $F_0(x) = \operatorname{argmin} \Sigma L(y_i, \gamma)$

3 Вычисление  $r_{i,m} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$

4 Обучение регрессии  $r_{jm}$  в цикле  $j = 1 \dots m$

5 В цикле  $j = 1 \dots m$  вычисление  $r_{jm} = \operatorname{argmin} \Sigma L(y_i, F_{m-1}(x_i) + \gamma)$

6  $F_m(x) = F_{m-1}(x) + v \Sigma \gamma_m I(x \in R_{jm})$



**Время  
практики**